Researcher

CILT/VOL **5**
SAYI/ISSUE **01**
YIL/YEAR **2025**

ANKARA BİLİM ÜNİVERSİTESİ

Researcher

CİLT/VOL **05** SAYI/ISSUE **01** YIL/YEAR **2025**

e-ISSN:2717-9494

Yayıncı / Publisher: Ankara Bilim Üniversitesi / Ankara Science University
Basım Tarihi / Date of Publication: Temmuz 2025/ July 2025
Yayın Türü / Publication Type: Uluslararası Süreli Yayın / International Periodical
İletişim Bilgileri / Contact Information: Maltepe Mahallesi Şehit Gönenç Caddesi No: 5 Çankaya Ankara
Web Sitesi / Website: https://dergipark.org.tr/tr/pub/researcher
E-posta / E-mail: researcher@ankarabilim.edu.tr

Researcher uluslararası, hakemli ve yılda iki sayı yayınlanan dergidir. İngilizce ve Türkçe dilindeki metinler kabul edilir.
The Researcher is a peer-reviewed, international journal publishing two issues a year. Manuscripts are accepted in English and Turkish languages.

ANKARA, TEMMUZ 2025 / JULY 2025

Önsöz

Yayın hayatına 2013 yılında başlamış olan "Researcher: Social Sciences Studies" (RSSS), 2020 Ağustos ayı itibariyle "Researcher" ismiyle Ankara Bilim Üniversitesi bünyesinde yayın hayatına devam etmektedir. Fen Bilimleri alanına katkıda bulunmayı hedefleyen özgün araştırma makalelerinin yayımlandığı bir dergidir. Dergi, özel sayılar dışında yılda iki kez yayımlanmaktadır.

Amaçları doğrultusunda dergimizin yayın odağında; Endüstri Mühendisliği, Bilgisayar Mühendisliği ve Elektrik Elektronik Mühendisliği alanları bulunmaktadır. Dergide yayımlanmak üzere gönderilen aday makaleler Türkçe ve İngilizce dillerinde yazılabilir. Dergiye gönderilen makalelerin daha önce başka bir dergide yayımlanmamış veya yayımlanmak üzere başka bir dergiye gönderilmemiş olması gerekmektedir. Bir makalenin dergide yayımlanabilmesi için en az iki hakem tarafından olumlu rapor verilmesi gerekir.

Değerlendirme sonucu kabul edilen çalışmalar sırasıyla; intihal kontrolünün yapılması, kaynakça düzenlemesi, gönderme ve atıf kontrolü, mizanpaj ve dizgisinin yapılması süreçlerinden geçer.

Researcher, Dergipark üzerinden bilimsel araştırmaların içeriğine anında açık erişim sağlamaktadır.

Researcher makale işlem ücreti (gönderme, değerlendirme veya basım ücreti) ve makalelere erişim için abonelik ücreti talep etmediği için ücretsiz yayın yapan dergi statüsündedir, Dergimiz herhangi bir kâr amacı gütmemekte ve hiçbir gelir kaynağı bulunmamaktadır.

Baş Editör

Prof. Dr. Hakan ÇAĞLAR

Editör

Doç. Dr. Yavuz Selim ÖZDEMİR

İçindekiler / Index

ANKARA SCIENCE UNIVERSITY, RESEARCHER
Vol. 5, No. 1, July 2025
e-ISSN: 2717-9494
Research Article/Araştırma Makalesi

A Prioritization Analysis for UAVs in Disaster Response

Melike ERDOĞAN[1], Ayşegül KOSAK[2,*]
[1]ANR Technologies, Istanbul
[2]Department of Industrial Engineering, Düzce University, 81620, Düzce; https://orcid.org/0009-0001-7725-1217
* Corresponding Author: aysegulkosak0@gmail.com

Received: 15 October 2024; Accepted: 10 July 2025

Reference/Atıf: M. Erdoğan ve A. Koşak, "A Prioritization Analysis for UAVs in Disaster Response", Researcher, c. 05, sy. 01, ss. 1–16.

Abstract

Many natural disasters happen in the world every year. Disasters make it difficult for people to reach their basic needs and can cause unpredictable loss of life. For this reason, emergency response and resource management are of critical importance in case of disaster. Since there is serious damage to the transportation infrastructure after the disaster, the roads, bridges and railway lines become unusable, making it difficult for the teams and relief materials to reach the disaster area by traditional methods, creating danger and increasing the loss of life. In cases where transportation is not possible, the fact that unmanned aerial vehicles (UAVs) increase accessibility to the disaster area creates a serious advantage in disaster times. UAVs reduce the possible environmental effects causing delays in the event of a disaster compared to transportation by traditional methods, can take part in dangerous conditions, can provide medicines and food supply, and can provide fast and safe transportation of needs to disaster victims. In this study, UAVs that can provide the medicine and food needed after a disaster are prioritized by using a multi-criteria decision-making (MCDM) approach with six main criteria and twenty-one sub-criteria determined. Seven UAV alternatives have been analyzed to use in disaster times primarily. Analytical Hierarchy Process (AHP) method has been adopted to weigh the criteria and Average Distance to Solution (EDAS) method has been used for the evaluation and prioritization of alternatives. This study is the first to use the EDAS method for the supply of medicine and food in disaster situations. In order to compare the results of the study, Complex Proportional Assessment (COPRAS) and Technique for Order Preference by Similarity to Ideal Solutions (TOPSIS) multi-criteria decision-making methods have been also utilized. Sensitivity analyses have been performed to determine the effect of criteria weights on the ranking of alternatives. The reliability and robustness of the results have been investigated through comparison and sensitivity analyses.

Keywords: disaster response, humanitarian logistics, multi-criteria decision making, unmanned aerial vehicles

1. Introduction

Disasters often cause people to leave their homes or have difficulty meeting their basic needs. With the prioritization of unmanned aerial vehicles (UAVs) for the supply of food and medicine, disaster victims will be able to reach their basic nutritional and health needs. The initial few hours after a disaster, when victims have the best chance of surviving, are referred to as the crucial phase. All human resources that are available after a disaster must be directed toward the search and rescue effort. Furthermore, coordinating rescue efforts and expeditiously evaluating the disaster's impact are critical components of response activities. [1]. In case of disaster, food and medicine supply has a critical role in protecting the lives of disaster victims, ensuring their health, and contributing to the rapid recovery of society. Providing this supply regularly, effectively and quickly is considered a vital element in post-disaster response. Health risks such as water pollution, food insecurity and hygiene problems may arise during disasters. UAVs play a critical role in drug supply to prevent the spread of diseases and intervene in existing health problems, and those who are less affected by environmental conditions can deliver drugs and other needs quickly. Because it is important to quickly access the disaster area and meet basic needs, UAVs provide support for medicine and food supplies, emergency aid and rescue operations. In this way, food and medicine supply will be provided in the fastest, safest and most efficient way in case of disaster [2]. Even if damaged and blocked roads pose obstacles in weather and land conditions, it enables unmanned aerial vehicles to quickly reach the disaster area and provide emergency food supply. This paper aimes to ensure the delivery of food and medicine in cases of emergency or disaster and to support

emergency aid operations. UAVs that can be used immediately after the disaster have been identified and a multi-criteria analysis has been proposed as to which of them should be used first. With the proposed UAV, it will be possible to complete its task quickly in food and medicine supply by being less affected by environmental conditions. This type of prioritization analysis is intended to evaluate the advantages and potential challenges of using UAVs in disaster situations. This will provide disaster response teams, governments and aid agencies with the ability to act more effectively and quickly. The main purpose of this analysis is to ensure the effective use of UAVs for food and medicine supply in disaster situations. With this evaluation, the effectiveness and durability of UAV models in ensuring the supply of food and medicine in case of disaster is analyzed with the criteria determined via experts and literature review. The fact that there are multiple alternatives and many evaluation criteria in the evaluation process makes it reasonable to conduct this analysis using MCDM approaches. For this purpose, firstly, the evaluation criteria were weighted with AHP, which is one of the most frequently used and reliable MCDM approaches, and then the UAV model recommended to be used as a priority among the UAV alternatives identified using the EDAS method was determined. In order to test the reliability and robustness of the study, comparative and sensitivity analyses were performed, and the results obtained under different scenarios were compared with the results obtained in the current analysis. In the following sections, the literature survey conducted for the study, the proposed hybrid methodology and the comparison analysis and sensitivity analysis are included

2. Literature Review

There are some studies in the literature in which UAVs are evaluated with MCDM methods for use in disaster situations. While analyzing these studies, a large number of research containing the keywords "disaster response", "multi-criteria decision making", "humanitarian logistics", "medicine and aid supply" and "unmanned aerial vehicle" were examined. However, a limited number of studies conducted for the purpose adopted in this paper were found. The studies are summarized as follows in Table 1 in terms of methods used, purpose, year and countries.

Table 1. Literature Research Results

	Author(s)	Year	Aim	Adopted Method(s)	Country
1	Gürbüz et al. [3]	2023	Answering the questions of how, where, in which way, with which species and what type of pesticides for farmers in agriculture in Kırıkkale Province	AHP, TOPSIS, (PROMETHEE) , (VIKOR)	Turkey
2	Kara et al. [4]	2023	Making the optimum choice among firefighter drones produced to transport liquids to intervene in fires	AHP and COPRAS	Turkey
3	Ecer et al. [5]	2023	Proposal for an integrated group decision-making framework to identify the best agricultural UAV	q-ROFNs, (LOPCOW), VIKOR	Turkey
4	Garg et al. [6]	2023	Developing sustainable drone delivery solutions	Systematic Literature Review (SLR)	USA
5	Silva et al. [7]	2023	Propose a model that can help decision-makers choose the most appropriate last-mile solution for historical centers	AHP and TOPSIS	Portugal
6	Banik et al. [8]	2023	Choosing the most suitable drone in different scenarios related to medical supply distribution	Graph Theory and Matrix Approximation (GTMA)	USA
7	Tesic et al. [9]	2023	Surveying flooded areas during floods and providing necessary supplies, food and water	Multi-Attribute Boundary Approach Area Comparison (DIBR-Rough Mabac)	Serbia
8	Hossain et al. [10]	2022	Estimating the overall performance of drone technology through four main criteria (factors)	Bayesian Network (BN) approach, Sensitivity Analysis	USA

9	Zahir H. et al. [11]	2022	Strategic framework proposal for optimizing drone capabilities in cities' disaster response	Participatory Action Research (PAR) Approach	Malaysia
10	Dukic et al.	2022	Solving the problem of more efficient and economical training of combat crews on short-range air defense systems	AHP and TOPSIS	Serbia
11	Aktas and Kabak [12]	2022	Proposing a model to determine the most suitable drone alternative	Pythagorean Fuzzy Weighted Aggregate Total Product Evaluation	Turkey
12	Kara et al. [13]	2022	Choosing a drone with similar features produced for material transportation	AHP, TOPSIS and PROMETHEE	Turkey
13	Özaslan et al.	2021	It was determined that there are no studies on the selection of piston single-engine aircraft in individual purchasing in Turkey.	AHP and TOPSIS	Turkey
14	Rejeb, et al. [14]	2021	Analyzing the potential applications of drones in the humanitarian field and structuring research on the subject	Humanitarian Logistics (HL) Research	USA
15	Sohaib Khan et al. [15]	2021	Selection of the highest-order drone by applying the TOPSIS method to the drones on the market at the desired cost	AHP and TOPSIS	Pakistan
16	Fu et al. [16]	2021	Examining the opinions of experts at the consultation stage before enacting a law on civilian UAVs	AHP and Triple Helix Model (THM)	Taiwan
17	Ergun et al. [17]	2021	Development of a game theoretical model for emergency logistics planning	Cooperative Game Theory	Turkey
18	Zhang et al.	2021	Proposing a customized model to identify the top three benefits of drones at both the personal and community level	Fuzzy Analytical Network Process (ANP), Fuzzy Decision Making Trial and Evaluation Laboratory (DEMATEL)	USA
19	Nur et al. [18]	2020	Comparing a range of existing last-mile delivery drones suggests an extensive list of criteria.	TOPSIS and AHP	USA
20	Glantz et al. [19]	2020	How UAV capabilities are used in disaster management, their current use in disaster management	Examination of MCDM Methods	USA
21	Slavika Dožić [20]	2019	It is to identify and classify the problems solved using the multi-criteria decision-making method in the aviation industry.	Examination of MCDM Methods	Serbia
22	Değirmen et al. [1]	2018	Route planning of unmanned aerial vehicles to be used in disaster areas	Clustering and Mathematical Programming	Turkey
23	Özdemir and Baslıgil [21]	2016	Examining the purchase of aircraft from a Turkish airline using fuzzy numbers	FAHP, FANP, Choquet Integral	Turkey

This study addressed the MCDM problem of prioritizing unmanned aerial vehicles for the use of emergency response teams. When we look at Figure 1, which was prepared based on the studies classified by country in Table 1, Turkey is one of the countries with the highest rate, with 39%, among the countries where studies on the subject have been conducted. After Türkiye, the USA is among the countries where the most studies are observed, with a rate of 31%. In the following order, Serbia with 13%, Portugal with 5%, Pakistan, Taiwan and Malaysia with 4% come next. Our country carries out many studies for the disaster situations that have been experienced and may be experienced, and with the intensity of studies for disaster situations, it is a pioneer for the studies in other countries.

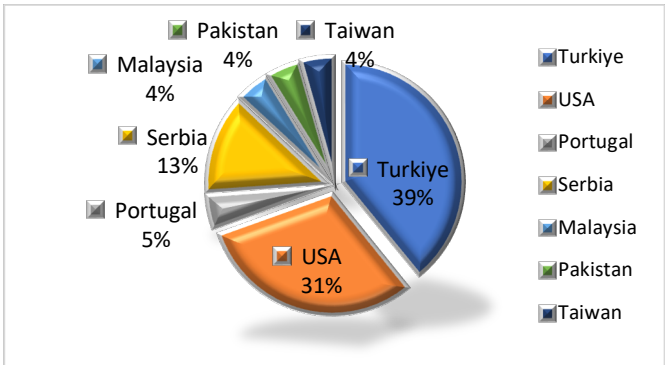

Figure 1. Studies according to their country

MCDM methods are frequently preferred to prioritize unmanned aerial vehicles for use in medicine and food supply in disaster situations. The most preferred methods are AHP, TOPSIS and PROMETHEE respectively. The distributions of these methods according to Table 1 are as in Figure 2. The AHP approach has been seen as the most frequently used MCDM method for UAV-related studies.

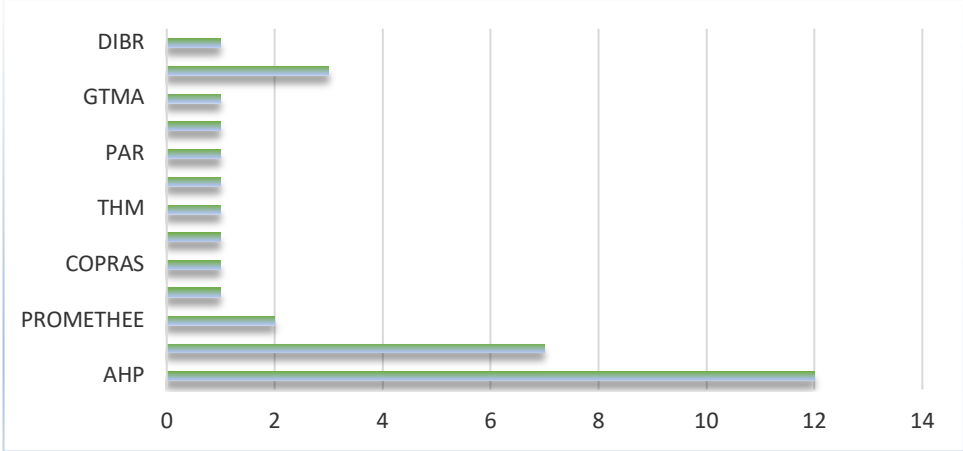

Figure 2. Studies according to their adopted methodology

3. Materials and Methods

In this study, a prioritization analysis has been conducted for the use of UAVs in food and medicine supply in case of post-disaster. Since there are multiple criteria and alternatives in the evaluation process, MCDM methods were adopted for the decision-making process. A hybrid AHP-EDAS MCDM methodology has been used in the study. The evaluation criteria have been weighted with the AHP method, and then the alternatives have been ranked with the EDAS method. With sensitivity analysis, the importance of criterion weights in ranking with different scenarios has been examined; comparative analysis has been carried out with TOPSIS and COPRAS methods and the results have been compared with the rankings obtained as a result of the EDAS method. The flowchart of the paper is shown in Figure 3.

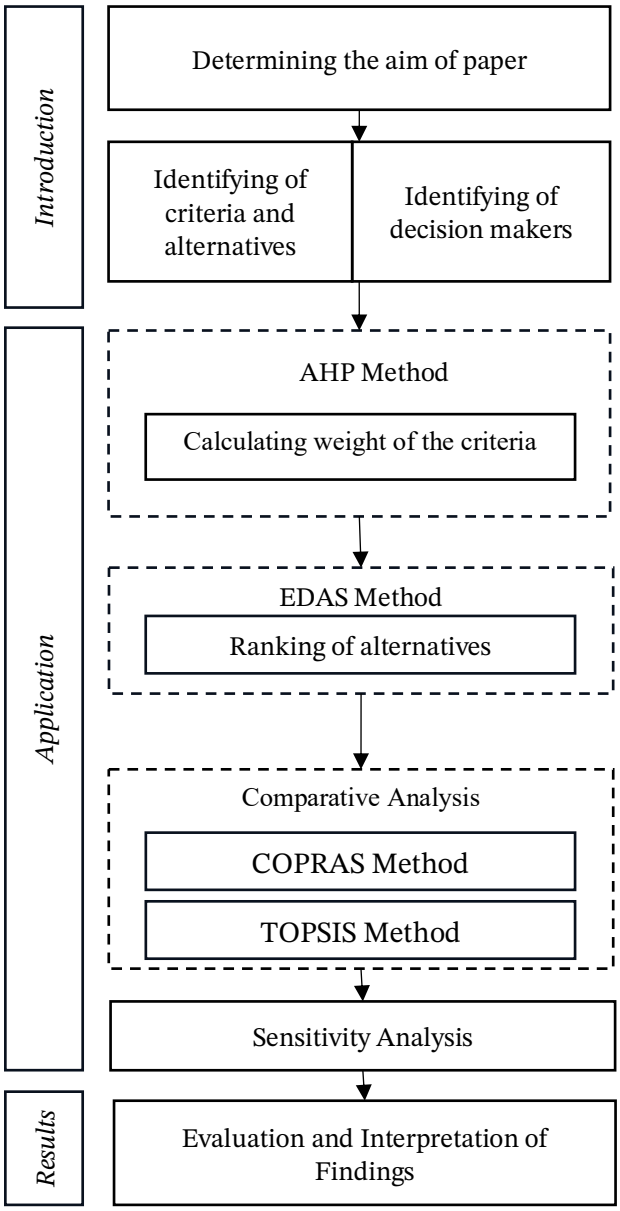

Figure 3. Flow chart of the proposed methodology

3.1 AHP Method

The AHP technique is a mathematical strategy that takes into account all priorities while making a decision. The most basic form employed in the AHP technique to structure a decision problem is a three-level hierarchy: the top-level decision aim, followed by a second level of criteria against which alternatives would be evaluated at the third level. The factors influencing the decision are ordered gradually. The structure's objective is to make it feasible to determine the importance of components at a specific level, based on some or all of the elements at the preceding level [22]. The scale to compare the elements to each other is used in the AHP method shown in Table 2.

Table 2. Linguistic Variables and Equivalents

Numerical Value	Definition	Explanation
1	Equally important	Two options are equally important
3	Moderately important	Moderately preferred one criterion over another
5	Strongly important	Experience and judgment have favored one criterion over another
7	Very strongly important	One criterion is considered superior to the other
9	Extremely important	Extremely more important and preferred
2,4,6,8	Intermediate values	It is used when compromise is necessary

The steps of the AHP method is given in the following [23]–[27];
Step 1: The decision problem is transformed into a hierarchical structure and a comparison matrix is created to show how the criteria are compared to each other as in Eq. (1). While creating the comparison matrix, experts can utilize the linguistic variables in Table 2.

$$B = [b]_{n\,x\,n} \tag{1}$$

Step 2: The consistency of each comparison matrix is checked. If the Consistency Ratio (CR) is less than 0.1, the matrix is considered consistent; otherwise, experts should evaluate the criteria again. Eq. (2) shows the Consistency Index (CI). Eq. (3) shows how the consistency ratio is calculated.

$$CI = \frac{\lambda_{max} - n}{(n-1)} \tag{2}$$

$$CR = \frac{\mathrm{CI}}{RI} \tag{3}$$

*R*I in Eq. (3) shows the randomness index. This value varies depending on the number of criteria. Table 3 gives the RI values according to the number of elements used in the problem.

Table 3. Randomness Index

n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

Step 3: The comparison matrix is normalized with Eq. (4). Then, the weight of each criterion is calculated with Eq. (5).

$$b'_{ij} = \frac{b_{ij}}{\sum_{i=1}^{n} b_{ij}} \tag{4}$$

$$w_j = \frac{\sum_{i=1}^{n} b'_{ij}}{\mathrm{n}} \tag{5}$$

3.2 EDAS Method

EDAS method, which can be translated into Turkish as "Evaluation Based on Average Solution Distance", is a new decision-making approach developed by Ghorabaee et al. [28] for the solution of MCDM problems. EDAS method compared with many MCDM methods, for its validity, and was successfully applied in solving many problems [29]. The EDAS method consists of 6 steps shown below [28]:

Step 1: The decision matrix (X) is constructed as shown below by Eq. (6). X_{ij} shows the performance value of *i*th alternative on *j*th criterion.

$$X = \left[X_{ij}\right]_{n\times m} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix} \tag{6}$$

Step 2: Calculate the average solution according to all criteria, as in

$$AV \;=\; [AVj]1xm \tag{7}$$

AV_j represents the average of the criterion and is determined with Eq. (8).

$$AV_j = \frac{\sum_{i=1}^{n} X_{ij}}{n} \tag{8}$$

Step 3: Positive distance from the mean (PDA) and negative distance from the mean (NDA) matrices are defined according to the criterion type (benefit and cost) as follows.

$$PDA = \left[PDA_{ij}\right]_{n\times m} \tag{9}$$

$$NDA = \left[NDA_{ij}\right]_{n\times m} \tag{10}$$

If the criterion is benefit-based, Eq.s (11) and (12) are adopted.

$$PDA_{ij} = \frac{max\left(0, \left(X_{ij} - AV_j\right)\right)}{AV_j} \tag{11}$$

$$NDA_{ij} = \frac{max\left(0, \left(AV_j - X_{ij}\right)\right)}{AV_j} \tag{12}$$

If the criterion is cost-based, Eq. (13) and (14) are adopted.

$$PDA_{ij} = \frac{max\left(0, \left(AV_j - X_{ij}\right)\right)}{AV_j} \tag{13}$$

$$NDA_{ij} = NDA_{ij} = \frac{max\left(0, \left(X_{ij} - AV_j\right)\right)\right)}{AV_j} \tag{14}$$

where PDA_{ij} and NDA_{ij} show the positive and negative distance of i th alternative from the average solution in terms of jth criterion, respectively.

Step 4: SPi and SNi values are found with the help of the following equations.

$$SP_i = \sum_{j=1}^{m} w_j PDA_{ij} \tag{15}$$

$$SN_i = \sum_{j=1}^{m} w_j NDA_{ij} \tag{16}$$

Step 5: Normalization procedure for SPi and SNi values of all alternatives is conducted with the help of the following equations.

$$NSP_i = \frac{SP_i}{\max_i (SP_i)} \tag{17}$$

$$NSN_i = 1 - \frac{SN_i}{\max_i (SN_i)} \tag{18}$$

Step 6. The appraisal score (AS) of each of the alternatives is calculated.

$$AS_i = \frac{1}{2}(NSP_i + NSN_i) \tag{19}$$

ASi in equation (2.3.14) must satisfy the equality $0 \leq ASi \leq 1$.
Step 7. Sort the alternatives according to the evaluation score (AS) results. The alternative with the highest score is considered the best alternative among the candidates.

4. Real Case Application

In this paper, seven UAVs we identified for use in medicine and food supply in case of disaster have been evaluated and prioritized. In order to evaluate the alternatives via the evaluations of the decision makers, six main and twenty-one sub-criteria have been determined as a result of the literature search. The main criteria for this study have been determined as "social", "environmental", "economic" and "technological" and sub-criteria have been also specified in the same way and placed under the appropriate main criteria. During the determination of alternatives, seven alternative UAVs produced for use in medicine and food supply in case of disaster have been considered. Sancak and Jackal-M used in the Kahramanmaraş earthquake; Matternet and Zipline's Zip, used to transport medical supplies during the Covid-19 outbreak; In last mile transportation, Flytrex, Wingcopter 198 and Foxtech Gaia unmanned aerial vehicles are nationally and internationally purchasable, easy to use and preferred tools for combating disasters. The determined drones are taken into account as an alternative by using them for hours under disaster conditions, shaping them according to the needs and considering the capabilities that can withstand difficult conditions. In addition, attention has been paid to the fact that the selected alternatives are physically and cognitively capable and have gained experience. After identifying the seven alternatives that produce drones for use in medicine and food supply in disaster situations, the decision hierarchy has been established. The criteria are determined via the literature review [3]–[5], [7], [8], [10], [14] . The hierarchy created for this paper is shown in Figure 4.

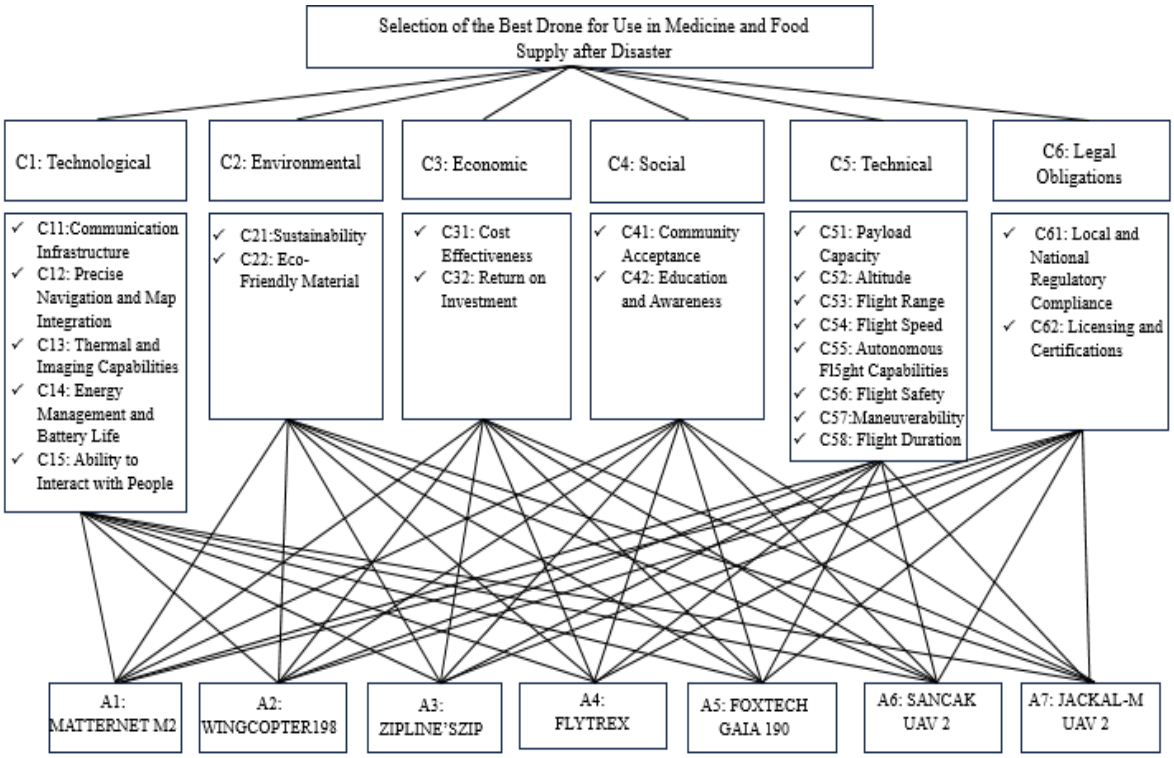

Figure 4. Hierarchical Representation of the Criteria, Sub-criteria and Alternatives

Criteria weights have been calculated using the data obtained from the evaluation of experts. Afterward, the criteria-alternative evaluations obtained by experts are quantified and the ranking of the alternative has been determined using the EDAS method. The results also compare with COPRAS and TOPSIS methods and sensitivity analysis in different scenarios have been conducted. A hierarchical structure has been created by determining criteria and alternatives in line with the selection of drones for use in medicine and food supply. The criteria determined for this decision problem have been evaluated by three experts according to their level of importance. The qualifications of the experts in this problem are presented in Table 4.

Table 4. Information about Experts

Experts	Experience	Job
E1	2023 Teknofest Combat UAV Competition/Composite UAV Production 2022-23 Teknofest Rocket Competition Finalist/ Rocket Production	Member of Düzce University Aviation and Space Technologies Community/Mechanical Engineer
E2	Composite Defense Industry UAV Design and Production	Member of Düzce University Aviation and Space Technologies Community/Mechanical Engineer
E3	2023 Composite UAV Design and Analysis/Ansys Analysis and Solidwork	Düzce University Career Community Chairman of the Board/Mechanical Engineer

For calculating the criteria weights, experts have first evaluated the main criteria to their degree of importance. Afterwards, the geometric mean of the expert evaluations was calculated for each main criterion. Table 5 shows the results of the aggregated evaluation of the main criteria comparison.

Table 5. Aggregated values of expert evaluations for main criteria weighting

	C1	**C2**	**C3**	**C4**	**C5**	**C6**
C1	1,000	0,382	0,315	0,369	2,000	3,557
C2	2,621	1,000	1,442	0,500	2,520	4,932
C3	3,175	0,694	1,000	0,550	3,557	4,718
C4	2,714	2,000	1,817	1,000	3,915	5,518
C5	0,281	0,203	0,212	0,181	1,000	0,397
C6	0,500	0,397	0,281	0,255	2,520	1,000

Similarly, for the sub-criteria sets of each main criterion, aggregated evaluations for expert opinions have been calculated with the geometric mean function. Meanwhile, all these pairwise comparison matrices have been checked for consistency. As a result of the consistency analysis, the consistency ratios of all pairwise comparison matrices have been found to be less than 0.1 and the weight calculation stages have been initiated. The consistency ratios calculated for each pairwise comparison matrix are presented in Table 6.

Table 6. Consistency Indexes for Pairwise Comparison Matrices

Comparison Matrix	**Consistency Ratio**
C1	0,0448
C2	0,0000
C3	0,0000
C4	0,0000
C5	0,0640
C6	0,0000

By applying the AHP steps presented in Section 3.1, the weights of the main criteria and sub-criteria were calculated. These weights are shown in Table 7.

Table 7. Weights of criteria

Main Criteria	**Main Criteria Weight**	**Sub-Criteria**	**Local Weights**	**Global Weights**
C1	0.1132	C11	0.4950	0.0561
		C12	0.2410	0.0272
		C13	0.1320	0.0150
		C14	0.0870	0.0099
		C15	0.0440	0.0050
C2	0.2230	C21	0.7565	0.1687
		C22	0.2435	0.0543
C3	0.2208	C31	0.8208	0.1812
		C32	0.1792	0.0396
C4	0.3235	C41	0.7965	0.2577
		C42	0.2035	0.0658
C5	0.0777	C51	0.2730	0.0212
		C52	0.2280	0.0177
		C53	0.1590	0.0123
		C54	0.1310	0.0102
		C55	0.0810	0.0063

		C56	0.0570	0.0045
		C57	0.0410	0.0032
		C58	0.0300	0.0024
C6	0.0418	C61	0.8445	0.0353
		C62	0.1555	0.0065

As a result of the multi-criteria analysis, the criterion with the highest weight has been determined as the "social" criterion with the degree of importance 0.325. The fact that the social criterion has the biggest weight for UAVs is acceptable and highlights the significance of increasing public knowledge. Additionally, the fact that these vehicles are designed to serve people alters their use and preference by providing an answer to the question of how social they are. The main criterion with the lowest criterion weight is "Legal and Regulatory" with a significance level of 0.041. The adverse conditions experienced in disaster situations bring many risks and require all precautions to be taken as soon as possible. Considering all these risks, it cannot be expected to meet the legal and regulatory criteria as a priority, therefore it can be reasonably accepted that the relevant criterion is in the last place.
After obtaining the criteria weights, the EDAS approach has been used to rank the determined alternatives. After the criteria weights had been obtained, the EDAS approach was used to rank the determined alternatives. Again, the steps in Section 3.2 have been processed on the criterion-alternative evaluation matrix provided by the decision makers and the importance rankings of the alternatives have been found. Table 8 shows the ranking results for the alternatives.

Table 8. Ranking of Alternatives with the EDAS Method

Alternative	*SP*	*NSP*	*SN*	*NSN*	AS_i	Ranking
A1	0.041	0.242	0.158	0.210	0.226	6
A2	0.107	0.624	0.104	0.483	0.554	5
A3	0.172	1.000	0.135	0.324	0.662	4
A4	0.120	0.700	0.059	0.705	0.703	2
A5	0.051	0.301	0.201	0.000	0.151	7
A6	0.126	0.736	0.012	0.935	0.836	1
A7	0.093	0.540	0.041	0.794	0.667	3

The first alternative that comes to mind is the Jackal-M UAV2 unmanned aerial vehicle. This UAV is one of the vehicles used in the Kahramanmaraş Earthquake in Türkiye on February 6. Jackal-M's communication infrastructure is supported by the Satcom satellite and has begun to be exported to the UK in the international market. This UAV has long hours of flight experience and has the highest flight speed, longest flight range and payload capacity compared to other UAVs. For these reasons, it is not surprising that this alternative is in the first place. Wingcopter 198 (A4) has become the second alternative after Jackal-M UAV2. Last in line was the Foxtech GAIA 190 alternative, which is generally used in last-mile logistics for food supply.

4.1. Comparison Analysis

In this section, comparisons of the results obtained with different MCDM methods are presented.

4.1.1. TOPSIS Method

The TOPSIS method is an MCDM method developed by Hwang and Yoon in 1981 and used in many decision-making problems [30]. The method was developed to make preference rankings based on the basic principle of proximity of decision points to the ideal solution [30], [31]. The steps of the TOPSIS method can be followed in [30]. The criteria alternative evaluations received from experts have been taken as input in the TOPSIS method and a ranking of the alternatives has been obtained again. Table 9 shows the importance rankings for the alternatives as a result of the conducting of the TOPSIS steps.

Table 9. Rankings by TOPSIS Method

Alternatives	S_i^-	S_i^*	C_i^*	**Ranking**
A1	0.03180	0.06695	0.32203	6
A2	0.06159	0.05082	0.54793	1
A3	0.05744	0.0617	0.48194	5
A4	0.05069	0.05266	0.49045	4
A5	0.02609	0.07642	0.25452	7
A6	0.04569	0.04154	0.52373	2
A7	0.04478	0.04464	0.50080	3

According to the prioritization analysis has been conducted with the TOPSIS method, it is seen that the ranking is found as A2>A6>A7>A4>A3>A1>A5. Alternative A2, with the highest Ci* value, has ranked first, and alternative A5, with the smallest degree of closeness, has ranked last.

4.1.2. COPRAS Method

Complex Proportional Assessment (COPRAS) method is an MCDM method that can evaluate qualitative and quantitative criteria. It has been applied in many areas to rank and evaluate alternatives. The most important feature that distinguishes the COPRAS method from other MCDM methods is to compare the options with each other and reveal as a percentage how much better or worse they are than other options [32], [33]. The steps of the method can be followed in [32]. The results according to the COPRAS method are shown in Table 10.

Table 10. Rankings for COPRAS Method

Alternative	P_i	Q_i	N_i	COPRAS
A1	0.126	0.126	0.792	6
A2	0.143	0.143	0.901	5
A3	0.148	0.148	0.9304	4
A4	0.152	0.152	0.9528	2
A5	0.122	0.122	0.764	7
A6	0.159	0.159	1.000	1
A7	0.150	0.150	0.944	3

Table 10. shows the results and rankings for the COPRAS method and it is seen that the best alternative is A6. It can be seen that the A5 alternative is in the last place. The order of the alternatives is A6>A4>A7>A3>A2>A1>A5.

As a result of the comparative analysis, the A6 alternative, which has come first in the current calculations, has been again ranked first in the COPRAS method and second in the TOPSIS method. The A5 alternative, which has been placed last, is again the last in both the TOPSIS and COPRAS methods. It has been observed that there is no change in the rankings of the A,1 A5 and A7 alternatives. As a result of the comparative analysis, it can be concluded that the results are reliable. Figure 5 shows all the rankings in a chart.

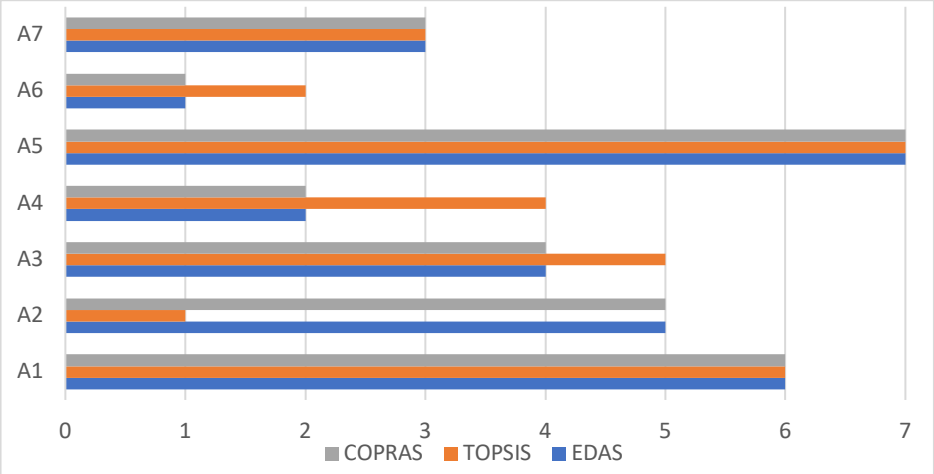

Figure 5. Results for comparison analysis

4.2. Sensitivitiy Analysis

To provide more insight into each selection or ranking methodology, performing sensitivity analysis against several of the assumptions of the base case inputs helps provide a more comprehensive view of the situation. When applying such methods, all sensitivities that provide maximum benefit to the factors driving the selection or ranking should be evaluated [34]. In this study, the criterion weights calculated within the scope of sensitivity analysis were changed and the effect of the change in weights on the results has been examined. Accordingly, the scenarios for the sensitivity analysis are;

Scenario 1 (S1): Interchange the weights of criteria K11, which has the highest criterion weight, and K58, which has the lowest criterion weight.

Scenario 2 (S2): Interchange the weights of criteria K11, which has the largest criterion weight, and K57, which has the next weight from the lowest criterion weight.

Scenario 3 (S3): Swapping the weights of criteria K11, which has the highest criterion weight, and K56, which has the 3rd lowest criterion weight.

Figure 6 shows the changing rankings according to the different scenarios mentioned.

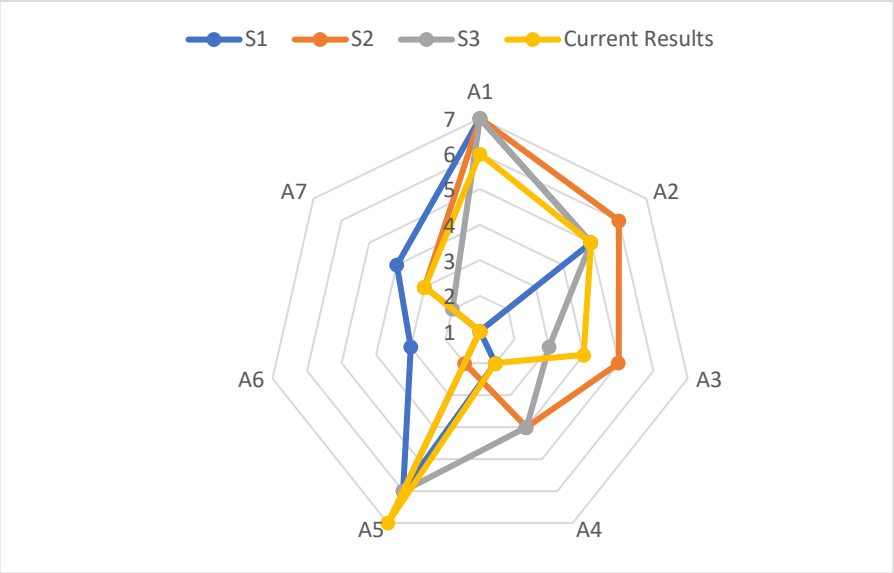

Figure 6. Sensitivity Analysis Results for Different Scenarios

When we examine the sensitivity analysis results, the alternative currently ranked first except for Scenario -1 has been again identified as the first place. Alternative A5, currently ranked the last

alternative, has a different ranking only in Scenario-2. Alternatives A3 and A7 have shown different performance rankings for different scenarios and can be stated to be very sensitive to changes in criterion weights. As a result of the sensitivity analysis, it can be claimed that the alternative rankings are affected by the change in the criteria weights, but the results are robust and reliable for our alternative, which is identified in the first place.

5. Conclusions and Future Suggestions

One of the most important topics recently encountered in the literature is the use of UAVs to provide food and medicine to disaster victims in order to save their lives, ensure their health, and thus contribute to the rapid recovery of society. UAVs help minimize the possibility of health risks such as water pollution, food insecurity and hygiene problems after disasters by enabling the timely supply of medicines. In this study, an MCDM analysis has been conducted to determine which UAVs should be used as a priority in order to rapidly supply food and medicine to the areas where they are needed in an emergency or disaster and to provide support for emergency relief operations. The UAVs that can be used for post-disaster food and medicine supply have been determined and a multi-criteria analysis has been proposed regarding which of these UAV alternatives should be used first. This prioritization analysis aims to evaluate the advantages and potential challenges of using UAVs in disaster situations. Seven UAV alternatives that can be used primarily for medicine and food supply in times of disaster have been determined. The criteria that can be utilized to evaluate these UAVs have been revealed and the AHP method has been adopted in weighting these criteria. The EDAS method has been used to evaluate and prioritize the alternatives. As a result of the multi-criteria analysis, the Jackal-M UAV2 alternative has been determined as the aircraft that can be used primarily in post-disaster medicine and food supply. In order to compare the results of the study, COPRAS and TOPSIS methods have been also adopted. Sensitivity analyses have been undertaken to examine the influence of criteria weights on alternative ranking. The reliability and robustness of the results have been explored using comparison and sensitivity studies. For future studies, fuzzy set theory can be adopted to consider the uncertainty in the decision-making process.

Contribution of Researchers

All authors contributed equally to the writing of this article.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethics committee approval (if needed)

No need to ethics committee approval statement.

References

[1] S. Değirmen, F. Çavdur, and A. Sebatlı, “Afet Operasyonları Yönetiminde İnsansız Hava Araçlarının Kullanımı: Gözetleme Operasyonları İçin Rota Planlama,” *Uludağ Üniversitesi Mühendislik Fakültesi Derg.*, vol. 23, no. 4, pp. 11–26, Dec. 2018, doi: 10.17482/UUMFD.455146.

[2] S. Firat and R. Dabak, “Afetlerde Yardım Malzemeleri Ulaştırmasında Yaşanan SOrunlar; İnsansız Hava Aracı Kullanımı,” *Meriç uluslararası Sos. ve Strat. Araştırmalar Derg.*, vol. 7, no. Özel Sayı, pp. 35–58, Oct. 2023, doi: 10.54707/MERIC.1325462.

[3] M. B. Gürbüz, Y. Erbaş, S. Ada, E. Karaşahin, E. Güven, and T. Eren, “Tarımda Drone Seçimi ve

Çizelgelemesi: Kırıkkale İli Örneği," *Int. Conf. Appl. Eng. Nat. Sci.*, vol. 1, no. 1, pp. 838–869, Jul. 2023, doi: 10.59287/ICAENS.1103.

[4] M. Kara, R. Yumuşak, and T. Eren, "Anız Yangınlarına Müdahale için İtfaiye Drone Seçimi: Giresun Örneği," *J. Aviat. Res.*, vol. 5, no. 1, pp. 1–15, Feb. 2023, doi: 10.51785/JAR.1180613.

[5] F. Ecer, İ. Y. Ögel, R. Krishankumar, and E. B. Tirkolaee, "The q-rung fuzzy LOPCOW-VIKOR model to assess the role of unmanned aerial vehicles for precision agriculture realization in the Agri-Food 4.0 era," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13373–13406, Nov. 2023, doi: 10.1007/S10462-023-10476-6/FIGURES/7.

[6] V. Garg, S. Niranjan, V. Prybutok, T. Pohlen, and D. Gligor, "Drones in last-mile delivery: A systematic review on Efficiency, Accessibility, and Sustainability," *Transp. Res. Part D Transp. Environ.*, vol. 123, p. 103831, Oct. 2023, doi: 10.1016/J.TRD.2023.103831.

[7] V. Silva, A. Amaral, and T. Fontes, "Towards sustainable last-mile logistics: A decision-making model for complex urban contexts," *Sustain. Cities Soc.*, vol. 96, p. 104665, Sep. 2023, doi: 10.1016/J.SCS.2023.104665.

[8] D. Banik, N. U. Ibne Hossain, K. Govindan, F. Nur, and K. Babski-Reeves, "A decision support model for selecting unmanned aerial vehicle for medical supplies: context of COVID-19 pandemic," *Int. J. Logist. Manag.*, vol. 34, no. 2, pp. 473–496, Mar. 2023, doi: 10.1108/IJLM-06-2021-0334/FULL/PDF.

[9] D. Z. Tešić, D. I. Božanić, and B. D. Miljković, "Application of MCDM DIBR-Rough Mabac Model for Selection of Drone for Use in Natural Disaster Caused by Flood," *Lect. Notes Networks Syst.*, vol. 659 LNNS, pp. 151–169, 2023, doi: 10.1007/978-3-031-29717-5_11.

[10] N. U. I. Hossain, N. Sakib, and K. Govindan, "Assessing the performance of unmanned aerial vehicle for logistics and transportation leveraging the Bayesian network approach," *Expert Syst. Appl.*, vol. 209, Dec. 2022, doi: 10.1016/J.ESWA.2022.118301.

[11] H. Zahir, M. S. Fathi, and A. F. Tharima, "Strategic framework of using drone in cities disaster response," in *The 9th AUN/SEED-Net Regional Conference on Natural Disaster (RCND 2021)*, 2021.

[12] A. Aktas and M. Kabak, "An Application of Interval Valued Pythagorean Fuzzy WASPAS Method for Drone Selection to Last Mile Delivery Operations," pp. 179–191, 2022, doi: 10.1007/978-3-030-98872-2_12.

[13] M. Kara *et al.*, "Acil Yardım Müdahalesi Yapan Birimler için Çok Ölçütlü Karar Verme Yöntemleri ile Kargo Drone Seçimi," *Türkiye İnsansız Hava Araçları Derg.*, vol. 4, no. 2, pp. 38–45, Dec. 2022, doi: 10.51534/TIHA.1148876.

[14] A. Rejeb, K. Rejeb, S. Simske, and H. Treiblmaier, "Humanitarian Drones: A Review and Research Agenda," *Internet of Things*, vol. 16, p. 100434, Dec. 2021, doi: 10.1016/J.IOT.2021.100434.

[15] M. Sohaib Khan, S. I. Ali Shah, A. Javed, N. Mumtaz Qadri, and N. Hussain, "Drone selection using multi-criteria decision-making methods," *Proc. 18th Int. Bhurban Conf. Appl. Sci. Technol. IBCAST 2021*, pp. 256–270, Jan. 2021, doi: 10.1109/IBCAST51254.2021.9393291.

[16] C. H. Fu, M. W. Tsao, L. P. Chi, and Z. Y. Zhuang, "On the Dominant Factors of Civilian-Use Drones: A Thorough Study and Analysis of Cross-Group Opinions Using a Triple Helix Model (THM) with the Analytic Hierarchy Process (AHP)," *Drones 2021, Vol. 5, Page 46*, vol. 5, no. 2, p. 46, May 2021, doi: 10.3390/DRONES5020046.

[17] S. Ergün, P. Usta, S. Z. Alparslan Gök, and G. W. Weber, "A game theoretical approach to emergency logistics planning in natural disasters," *Ann. Oper. Res.*, pp. 1–14, May 2021, doi: 10.1007/S10479-021-04099-9/FIGURES/4.

[18] F. Nur, A. Alrahahleh, R. Burch, K. Babski-Reeves, and M. Marufuzzaman, "Last mile delivery drone selection and evaluation using the interval-valued inferential fuzzy TOPSIS," *J. Comput. Des. Eng.*, vol. 7, no. 4, pp. 397–411, Aug. 2020, doi: 10.1093/JCDE/QWAA033.

[19] E. J. Glantz, "UAV Use in Disaster Management.," pp. 914–921, 2020.

[20] S. Dožić, "Multi-criteria decision making methods: Application in the aviation industry," *J. Air Transp. Manag.*, vol. 79, p. 101683, Aug. 2019, doi: 10.1016/J.JAIRTRAMAN.2019.101683.

[21] Y. Ozdemir and H. Basligil, "Aircraft selection using fuzzy ANP and the generalized choquet integral method: The Turkish airlines case," *J. Intell. Fuzzy Syst.*, vol. 31, no. 1, pp. 589–600, Jun. 2016, doi: 10.3233/IFS-162172.

[22] E. Atmaca, E. Aktaş, and H. N. Öztürk, "Evaluated Post-Disaster and Emergency Assembly Areas Using

Multi-Criteria Decision-Making Techniques: A Case Study of Turkey," *Sustain. 2023, Vol. 15, Page 8350*, vol. 15, no. 10, p. 8350, May 2023, doi: 10.3390/SU15108350.

[23] T. L. Saaty, "Decision-making with the AHP: Why is the principal eigenvector necessary," *Eur. J. Oper. Res.*, vol. 145, no. 1, pp. 85–91, Feb. 2003, doi: 10.1016/S0377-2217(02)00227-8.

[24] T. L. Saaty, "Decision making with the Analytic Hierarchy Process," *Int. J. Serv. Sci.*, vol. 1, no. 1, 2008, doi: 10.1504/IJSSCI.2008.017590.

[25] T. L. Saaty, *Fundamentals of decision making and priority theory with the analytic hierarchy process*, vol. 6. RWS publications, 2000.

[26] T. L. Saaty, "A scaling method for priorities in hierarchical structures," *J. Math. Psychol.*, vol. 15, no. 3, pp. 234–281, Jun. 1977, doi: 10.1016/0022-2496(77)90033-5.

[27] A. Ulutaş and D. Çelik, "Transpalet Seçimi Probleminin AHP VE EDAS Yöntemleri ile Değerlendirilmesi," *Bus. Manag. Stud. AN Int. J.*, vol. 7, no. 2, 2019.

[28] M. K. Ghorabaee, E. K. Zavadskas, L. Olfat, and Z. Turskis, "Multi-Criteria Inventory Classification Using a New Method of Evaluation Based on Distance from Average Solution (EDAS)," *Informatica*, vol. 26, no. 3, pp. 435–451, Jan. 2015, doi: 10.15388/INFORMATICA.2015.57.

[29] A. Özbek and M. Engür, "Çok Kriterli Karar Verme Yöntemleriyle Öğrenci İşleri Otomasyon Seçimi," *Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilim. Fakültesi Derg.*, vol. 21, no. 1, pp. 1–18, Jun. 2019, doi: 10.33707/AKUIIBFD.515581.

[30] C.-L. Hwang and K. Yoon, *Multiple Attributes Decision Making Methods And Applications*. Berlin: Springer, 1981.

[31] A. Karaşan, İ. Kaya, and M. Erdoğan, "Location selection of electric vehicles charging stations by using a fuzzy MCDM method: a case study in Turkey," *Neural Comput. Appl.*, 2018, doi: 10.1007/s00521-018-3752-2.

[32] E. Kazimieras Zavadskas, A. Kaklauskas, T. Vilutiene, E. Kazimieras ZAVADSKAS, and T. Vilutienė, "Multicriteria evaluation of apartment blocks maintenance contractors: Lithuanian case study," *Int. J. Strateg. Prop. Manag.*, vol. 13, no. 4, pp. 319–338, 2009, doi: 10.3846/1648-715X.2009.13.319-338.

[33] S. B. Patil, T. A. Patole, R. S. Jadhav, S. S. Suryawanshi, and S. J. Raykar, "Complex Proportional Assessment (COPRAS) based Multiple-Criteria Decision Making (MCDM) paradigm for hard turning process parameters," *Mater. Today Proc.*, vol. 59, pp. 835–840, Jan. 2022, doi: 10.1016/J.MATPR.2022.01.142.

[34] Y. Tedarik *et al.*, "Yeşil Tedarik Zinciri Yönetiminde Çok Kriterli Karar Verme: Otomotiv Ana Sanayi Örneği," *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilim. Derg.*, vol. 14, no. 3, pp. 675–698, Dec. 2019, doi: 10.17153/OGUIIBF.491356.

ANKARA SCIENCE UNIVERSITY, RESEARCHER
Vol. 5, No. 1, July 2025
e-ISSN: 2717-9494
Research Article/Araştırma Makalesi

A New Multiobjective Harris Hawk Optimization Algorithm for the Diagnosis of Breast Cancer

Alara SERMUTLU[1], Tansel DÖKEROĞLU[2*]
[1]TED University, Computer Engineering Department, https://orcid.org/0009-0005-3479-4640
[2]TED University, Computer Engineering Department, https://orcid.org/0000-0003-1665-5928
* tansel.dokeroglu@tedu.edu.tr

Received: 14 January 2025; Accepted: 29 July 2025

Reference/Atıf: A. Sermutlu ve T. Dökeroğlu, "A new multiobjective Harris Hawk Optimization algorithm for the diagnosis of breast cancer", Researcher, c. 05, sy. 01, ss. 17–25.

Abstract

Breast cancer, a highly prevalent and life-threatening disease, affects millions of individuals worldwide, particularly women. Feature-based methods are widely employed for early diagnosis of breast cancer, and selecting the optimal feature set remains a significant and challenging problem. In this study, we introduce a novel Multi-objective Harris Hawk Optimization algorithm, which integrates an adaptive K-Nearest Neighbor classifier. Comprehensive experiments were conducted on two well-known datasets. The proposed approach achieves 31-45% reductions in the total number of selected features across all datasets, significantly lowering computational costs and improving the accuracy of diagnostics up to 95-97%.

Keywords: Breast cancer, Harris Hawk, metaheuristic, K Nearest Neighbor

1. Introduction

Multi-objective metaheuristic algorithms are effective in feature selection by optimizing both accuracy and the number of selected features simultaneously [1]. They aim to balance the trade-off between maximizing classification performance and minimizing feature subsets. By exploring Pareto-optimal solutions, these algorithms enhance model efficiency, reduce computational complexity, and avoid overfitting, making them suitable for high-dimensional datasets in various applications. The Harris Hawk Optimization (HHO) algorithm is a recent nature-inspired metaheuristic optimization technique introduced in 2019 [2]. It mimics the cooperative hunting behaviour of Harris hawks in the wild, where they target prey in groups using both surprise and persistence strategies. HHO is designed to solve complex optimization problems by balancing exploration and exploitation in the search space.

In a recent study, Dokeroglu et al. proposed a multiobjective HHO for binary classification, aiming to reduce selected features while maximizing prediction accuracy. The algorithm demonstrates superior performance on benchmark datasets and a COVID-19 dataset [3]. Piri, J., & Mohapatra proposed a multi-objective optimization problem by proposing a Multi-Objective Quadratic Binary HHO (MOQBHHO) technique, integrating KNN as a wrapper classifier and crowding distance for solution selection [4]. Experimental results on twelve medical datasets demonstrate that MOQBHHO outperforms deep-based FS methods and other multi-objective algorithms in achieving an optimal trade-off between feature selection and classification accuracy. Selim et al. introduced an improved HHO algorithm for optimal Distributed Generation placement in radial distribution systems, aiming to minimize power loss, reduce voltage deviation, and improve voltage stability [5]. By enhancing HHO with a rabbit location mechanism and employing grey relation analysis for Pareto solutions, the proposed methods demonstrate superior performance on IEEE 33-bus and 69-bus systems compared to other optimization techniques.

Thawkar proposed a hybrid CSAHHO algorithm, combining the Crow Search Algorithm (CSA) and HHO, for feature selection and classification of masses in mammograms [6]. Using ANN and SVM classifiers, CSAHHO achieves superior performance with 97.85% accuracy, outperforming original CSA, HHO, and other state-of-the-art algorithms while using fewer features to enhance breast cancer diagnosis. Bandyopadhyay et al. proposed a two-stage pipeline for COVID-19 detection in CT scans, combining feature extraction with DenseNet and feature selection using an HHO algorithm enhanced

with Simulated Annealing (SA) and Chaotic initialization [7]. Evaluated on the SARS-COV-2 CT-Scan dataset, the method achieves an accuracy of 98.85% and reduces selected features by 75%, outperforming many state-of-the-art and hybrid meta-heuristic algorithms in both accuracy and feature reduction.

In this study, we implemented a multi-objective HHO algorithm for feature selection on the Wisconsin breast cancer dataset (WBCD) [8] and Wisconsin diagnostic breast cancer (WDBC) dataset, achieving notable results. The algorithm successfully balanced the trade-off between classification accuracy and feature reduction, improving accuracy significantly while reducing the number of selected features by 31-45%. This reduction enhances model efficiency and decreases computational complexity without compromising performance. By exploring Pareto-optimal solutions, the proposed approach demonstrates its ability to identify relevant features effectively. These results highlight the algorithm's potential for handling high-dimensional datasets, providing a robust method for feature selection in various classification tasks. Its performance surpasses many existing techniques.

2. Poroposed Multiobjective HNO Algorithm

This section briefly explains the proposed multiobjective HHO algorithm for the diagnosis of breast cancer. The algorithm begins by initializing a population of hawks, each representing a candidate solution. During the optimization process, the hawks adopt different strategies to simulate their natural hunting behaviours. In the exploration phase, hawks search for prey by randomly moving across the search space, promoting diversity and avoiding premature convergence. In the exploitation phase, hawks converge towards the prey using dynamic strategies, such as surprise pounce and soft or hard besiege tactics, to refine solutions and exploit the best regions.

One of HHO’s strengths is its simplicity and adaptability, making it suitable for various optimization problems in engineering, feature selection, scheduling, and machine learning. It is computationally efficient and requires fewer parameters compared to many other metaheuristics. Researchers have further enhanced HHO with hybrid models and variants, improving its performance in specific applications. Overall, HHO has proven to be a robust and versatile algorithm for solving real-world optimization challenges.

The HHO algorithm incorporates key parameters to mimic the hawks' energy and hunting behaviors, enhancing its ability to balance exploration and exploitation effectively. Among these parameters, energy (E) and r play crucial roles in determining the hawks' strategies:

The energy level of the prey (E) is modeled as a dynamic parameter to simulate the prey's attempt to escape. It decreases linearly over iterations and is represented as: $E=2E0\ (1-t/T)$, where E0 is the initial energy, t is the current iteration, and T is the maximum number of iterations.

When $|E|>1$, the Hawks focus on exploration, searching widely across the search space. When $|E|\leq 1$, the hawks shift to exploitation, zeroing in on the prey with more refined strategies like "soft besiege" or "hard besiege."

The parameter random factor (r) is a random value in the range [0, 1], used to probabilistically decide the Hawks’ movement strategy: $r < 0.5$: Indicates that hawks move randomly, mimicking the unpredictability of nature. r \geq 0.5: The hawks target the prey directly, focusing on convergence.

These parameters enable HHO to dynamically adjust its behaviour, making it versatile for a wide range of optimization problems. By modulating energy and incorporating randomness, the algorithm balances exploring the search space and exploiting promising regions to find optimal solutions efficiently. Figure 1 shows the energy level (E), q and r values of the HHO metaheuristic according to its diversification and intensification efforts. Algorithm 1 presents the details of the proposed multiobjective HHO algorithm for the diagnosis of breast cancer.

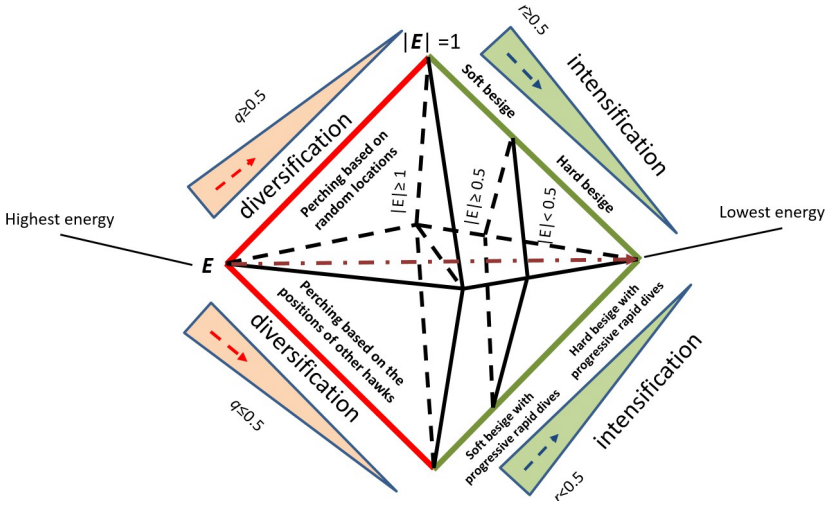

Figure 1. The steps of the HHO metaheuristic according to the energy level (E), q and r values.

Algorithm I. Multiobjective Harris Hawk Optimization Algorithm for feature selection

1	Initialize the population of hawks randomly (binary representation for feature subsets)	15	if r ≥ 0.5 and |E| ≥ 0.5 then Perform soft besiege
2	Set algorithm parameters (max iterations, population size, etc.)	16	else if r ≥ 0.5 and |E| < 0.5 then Perform hard besiege
3	Evaluate each hawk using two objectives:	17	else if r < 0.5 and |E| ≥ 0.5 then Perform soft besiege with rapid dives
4	1. Classification accuracy of the selected features (maximize)	18	else Perform hard besiege with rapid dives end if
5	2. Number of selected features (minimize)	19	end if Apply a repair mechanism if a solution violates constraints (e.g., empty feature subset)
6	Use a dominance-based mechanism (e.g., Pareto dominance) to identify the best solutions (prey)	20	end for Evaluate new solutions and update Pareto front
7	for each iteration do:	21	Update prey (best non-dominated solutions) if better solutions are found
8	Update to escaope energy of the prey(E)	22	end for Return the Pareto front of solutions (trade-off between accuracy and number of features)
9	for each hawk do:		
10	if |E| ≥ 1 then		
11	Perform exploration:		
12	Update hawk's position using a random search in binary space		
13	Else		
14	Perform exploitation:		

3. Experimental setup and evaluation of the results

The experiments in this study were conducted on a Huawei MateBook 14 laptop equipped with an AMD Ryzen 7 4800H processor, 16 GB of RAM, and 512 GB SSD, running the Windows 10 operating system.

Two datasets are used during our experiments, Wisconsin breast cancer dataset (WBCD) [8] and Wisconsin diagnostic breast cancer dataset (WDBC) [9]. The WBCD dataset consists of 699 samples characterized by nine numerical features. These features, derived from Fine Needle Aspiration (FNA) samples, capture various cellular and structural characteristics like cell size, shape and mitoses. The

dataset contains 16 incomplete records, which were excluded during preprocessing to ensure data homogeneity. This resulted in 683 complete samples, distributed as 444 benign and 239 malignant cases. The WDBC dataset includes 569 instances, each with 30 features extracted from a digitized image of an FNA of a breast mass. The features in the image represent the characteristics of the cell nuclei [9]. The details of the datasets are presented in Table 1.

Table 1. The details of the breast cancer datasets used in the experiments.

Dataset	Features	Instances	Malignant	Benign
WBCD	9	683	239 (35%)	444 (65%)
WDBC	30	569	212 (37.2%)	357 (62.8%)

In this part of our study, we conducted experiments on our datasets to evaluate the algorithm's performance with different numbers of generations and determine the optimal value for this parameter. A population size of ten individuals was chosen for this study, as it was previously shown to be sufficient for achieving reliable optimization. We tested both WDBC and WBCD datasets with 1, 2, 5, 10, 20, 30, and 50 generations. For the WBCD dataset, the optimal number of generations was 30, with accuracy declining beyond this point. Whereas, for the WDBC dataset, the highest accuracy was achieved at 20 generations, after which performance dropped. These results indicate that the ideal number of generations varies by dataset, with 30 being optimal for WBCD and 20 for WDBC.

Figures 2 and 3 present the average accuracy levels for different numbers of generations tested on the WBCD and WDBC datasets respectively.

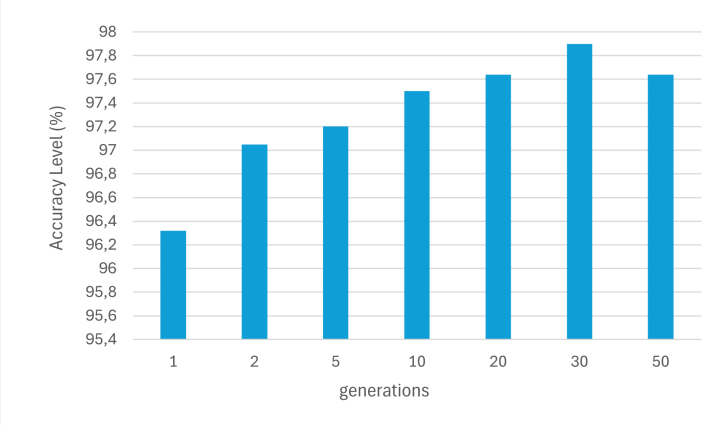

Figure 2. The average accuracy levels for different numbers of generations tested on the WBCD dataset.

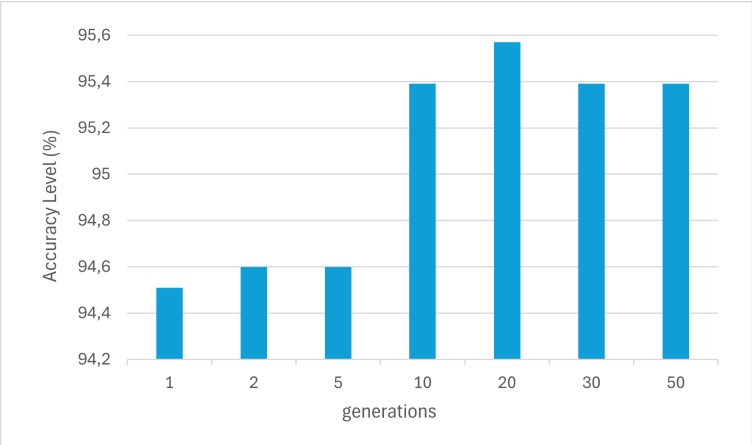

Figure 3. The average accuracy levels for different numbers of generations tested on the WDBC dataset.

The experiments, conducted over 10 iterations, are summarized in Table 2. For the WBCD dataset, the feature count was reduced from 9 to an average of 6.2, representing a 31% reduction, while maintaining an average accuracy of 97.0%. The maximum accuracy observed for the WBCD dataset was 97.5%. Similarly, for the WDBC dataset, the feature count was reduced from 30 to an average of 16.3 features, corresponding to a 45.6% reduction, with an average accuracy of 95.0%. The maximum accuracy for the WDBC dataset was 95.58%. These findings demonstrate the effectiveness of the proposed feature reduction approach in maintaining high classification accuracy.

Table 2. The average number of features in the populations and the accuracy levels for both datasets after executing 10 generations.

generations	1	2	3	4	5	6	7	8	9	10	**avg.**
WBCD # features	4	6	6	5	6	7	7	7	6	8	**6.2**
WBCD accuracy	96.7	97.35	96.76	97.35	96.91	96.91	97.06	97.06	97.35	97.5	**97.1**
WDBC # features	16	17	16	15	15	18	15	18	17	16	**16.3**
WBCD accuracy	95.0	95.22	94.87	94.69	94.04	94.87	95.58	95.04	95.04	95.58	**95.0**

In this section, we compared the accuracy levels and the number of features between the initial and final hawk populations to understand the impact of the evolutionary process on model performance. As shown in Figures 4 and 5, the initial population exhibited a wide range of accuracy levels and feature counts, indicating significant variability in the initial set of models. In contrast, the final population demonstrated a narrower range of accuracy levels with a higher, but more consistent feature count, around 7-8 features. This suggests that the algorithm effectively selected hawks with an optimal performance leading to improved consistency and potentially higher accuracy in the final population.

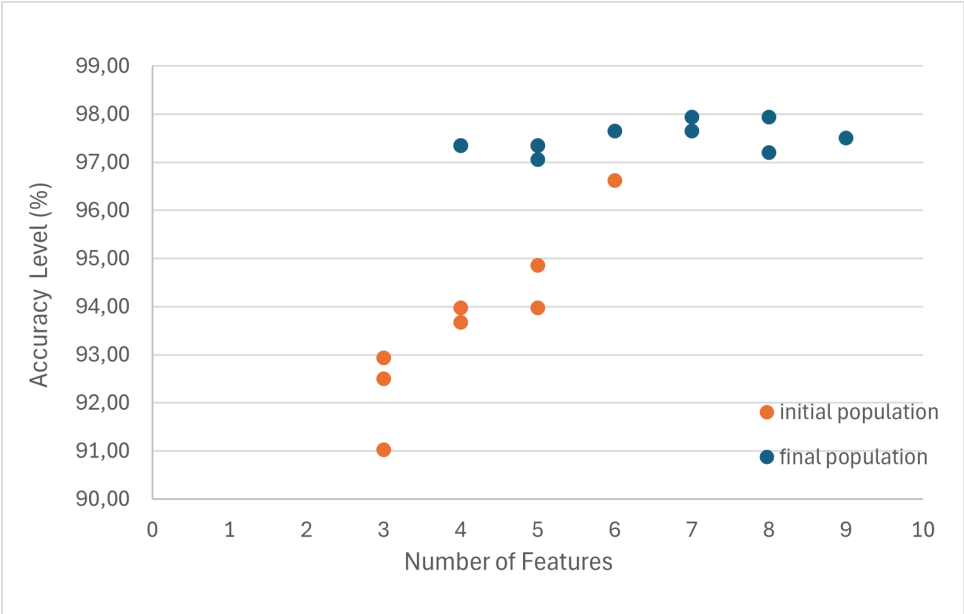

Figure 4. The initial and evolving populations of the proposed algorithm for the WBCD dataset through the generations.

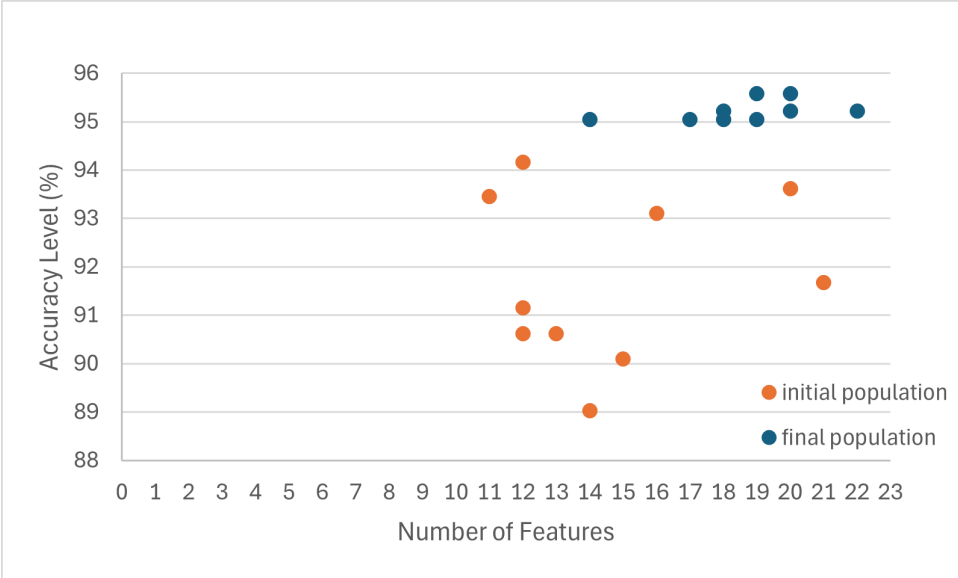

Figure 5. The initial and evolving populations of the proposed algorithm for the WDBC dataset through the generations.

4.Related Works

Several algorithms have been explored for breast cancer classification using the WDBC and WBCD datasets. In 2011, Doddipalli et al. reported high accuracy rates with decision tree classifiers like CART, achieving 96.99% accuracy on the WBCD dataset and 94.72% on the WDBC dataset [10]. Following this, Salama et al. (2012) demonstrated promising results with ensemble methods, achieving 97.28% accuracy on the WBCD dataset by fusing SMO, IBK, NB, and J48 classifiers, while individually, SMO achieved 97.72% on the WDBC dataset [11]. Furthermore, Aalaei et al. (2016) investigated the impact of GA-based feature selection, achieving 97.3% accuracy on the WDBC dataset with ANN and 96.9% on the WBCD dataset with PS-classifier [12]. Optimization techniques have also been explored, such as the adjusted Bat Algorithm (ABA) used by Tuba et al. (2016) to optimize SVM parameters, yielding accuracies of 96.99% on the WBCD dataset and 96.49% on the WDBC dataset [13]. Hybrid models,

like the HECFNN proposed by Alkhasawneh et al. (2018), combining CFNN and ENN, achieved 97.7% accuracy on the WBCD dataset [14].

Notable studies also include Mushtaq et al. (2019), who achieved 99.42% accuracy with KNN on the WBC dataset using Chi-square-based feature selection and the Manhattan distance function [15]. Singh et al. (2020) explored hybrid optimization approaches, combining GWO and WOA for SVM hyperparameter tuning, achieving 97.72% accuracy [16]. Wang et al. (2020) developed the IRFRE method, integrating Random Forest-based rule extraction with a multi-objective evolutionary algorithm, achieving 96.44% ± 3.76% accuracy on the WBCD dataset [17]. More recently, Badr et al. (2022) examined a hybrid GWO-SVM model, achieving 98.60% accuracy on the WDBC dataset with normalization scaling and 99.30% with their proposed scaling techniques [18]. These studies collectively demonstrate the potential of machine learning techniques in accurately classifying breast cancer, providing a foundation for further research and development in this area.
These studies demonstrate the potential of machine learning techniques in accurately classifying breast cancer, providing a foundation for further research and development in this area. The accuracy levels for WBCD and WDBC datasets for each study are summarized in Tables 3 and 4, respectively. It is important to note that, other than CART and the proposed MHHO algorithm, none of the reviewed studies explicitly reported the number of features used in their models.

Table 3. Comparison of the performance of the breast cancer classification methods on the WBCD Dataset

WBCD dataset			
Method	Feature #	Accuracy	Year
CART [10]	9	96.99	2011
SMO, IBK, NB, and J48 [11]	-	97.28	2012
ANN with FS [12]	-	96.90	2016
ABA-SVM [13]	-	96.99	2016
CFNN [14]	-	97.70	2018
KNN [15]	-	99.42	2019
IFRE[17]	-	96.44	2020
MHHO (our model)	**8**	**97.50**	**2025**

Table 4. Comparison of the performance of the breast cancer classification methods on the WDBC Dataset

WBCD dataset			
Method	number of features	Accuracy	Year
CART [10]	8	94.72	2011
SMO [11]	-	97.72	2012
PS [12]	-	97.30	2016
ABA-SVM [13]	-	96.49	2016
GWO-SVM [16]	-	99.30	2020
GWWOA-SVM [18]	-	97.72	2022
MHHO (our model)	**15**	**95.58**	**2025**

5. Conclusion and Future Work

Our study introduces a novel application of the HHO metaheuristic algorithm for multi-objective feature selection, integrated with an adaptive KNN classifier, for breast cancer diagnosis. Extensive experiments demonstrate superior accuracy in datasets and achieve a 31-45% reduction in the number of features significantly lowering computational costs. This feature reduction is accompanied by improved accuracy, confirming the efficiency of our approach. Overall, the proposed HHO algorithm provides a practical and effective solution for feature selection and classification in breast cancer diagnosis, offering promising results for future research. In our future work, we intend to study parallel versions of our proposed algorithm for the multiobjective HHO algorithm on GPU architectures.

Contribution of Researchers

All authors contributed equally to the writing of this article.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethics committee approval (if needed)

No need to ethics committee approval statement.

References

[1] Liu, Q., Li, X., Liu, H., & Guo, Z. (2020). Multi-objective metaheuristics for discrete optimization problems: A review of the state-of-the-art. Applied Soft Computing, 93, 106382.

[2] Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris Hawks optimization: Algorithm and applications. Future generation computer systems, 97, 849-872.

[3] Dokeroglu, T., Deniz, A., & Kiziloz, H. E. (2021). A robust multiobjective Harris' Hawks Optimization algorithm for the binary classification problem. Knowledge-Based Systems, 227, 107219.

[4] Piri, J., & Mohapatra, P. (2021). An analytical study of modified multi-objective Harris Hawk Optimizer towards medical data feature selection. Computers in Biology and Medicine, 135, 104558.

[5] Selim, A., Kamel, S., Alghamdi, A. S., & Jurado, F. (2020). Optimal placement of DGs in distribution system using an improved harris hawks optimizer based on single-and multi-objective approaches. IEEE Access, 8, 52815-52829.

[6] Thawkar, S. (2022). Feature selection and classification in mammography using hybrid crow search algorithm with Harris Hawks optimization. Biocybernetics and Biomedical Engineering, 42(4), 1094-1111.

[7] Bandyopadhyay, R., Basu, A., Cuevas, E., & Sarkar, R. (2021). Harris Hawks optimisation with Simulated Annealing as a deep feature selection method for screening of COVID-19 CT-scans. Applied Soft Computing, 111, 107698.

[8] Wolberg, W. (1990). Breast Cancer Wisconsin (Original) [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5HP4Z.

[9] Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1993). Breast Cancer Wisconsin (Diagnostic) [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5DW2B.

[10] Doddipalli, Lavanya & Rani, K.. (2011). Analysis of feature selection with classification: Breast cancer datasets. Indian Journal of Computer Science and Engineering (IJCSE). 2. 756-763.

[11] Salama, G. I., Abdelhalim, M. B., & Zeid, M. A. (2012). Breast cancer diagnosis on three different datasets using multi-classifiers. *International Journal of Computer and Information Technology, 1*(1), 1-10.

[12] Aalaei S, Shahraki H, Rowhanimanesh A, Eslami S. Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets. Iran J Basic Med Sci. 2016 May;19(5):476-82. PMID: 27403253; PMCID: PMC4923467.

[13] E. Tuba, M. Tuba and D. Simian, "Adjusted bat algorithm for tuning of support vector machine parameters," 2016 IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, Canada, 2016, pp. 2225-2232, doi: 10.1109/CEC.2016.7744063.

[14] Alkhasawneh, M. S., & Tay, L. T. (2018). A Hybrid Intelligent System Integrating the Cascade Forward Neural Network with Elman Neural Network. Arabian Journal for Science and Engineering, 43(12), 6737–6749. https://doi.org/10.1007/s13369-017- 2833-3

[15] Mushtaq, Z., Yaqub, A., Sani, S., & Khalid, A. (2019). Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets. Journal of the Chinese Institute of Engineers, 43(1), 80–92. https://doi.org/10.1080/02533839.2019.1676658

[16] I. Singh, R. Bansal, A. Gupta and A. Singh, "A Hybrid Grey Wolf-Whale Optimization Algorithm for Optimizing SVM in Breast Cancer Diagnosis," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), Waknaghat, India, 2020, pp. 286-290, doi: 10.1109/PDGC50313.2020.9315816.

[17] Wang, S., Wang, Y., Wang, D., Yin, Y., Wang, Y., & Jin, Y. (2020). An improved random forest-based rule extraction method for breast cancer diagnosis. Applied Soft Computing, 86, 105941. https://doi.org/10.1016/j.asoc.2019.105941

[18] Badr, E., Almotairi, S., Abdul Salam, M., & Ahmed, H. (2022). New sequential and parallel support vector machine with grey wolf optimizer for breast cancer diagnosis. Alexandria Engineering Journal, 61(3), 2520-2534. https://doi.org/10.1016/j.aej.2021.07.024

ANKARA SCIENCE UNIVERSITY, RESEARCHER
Vol. 5, No. 1, July 2025
e-ISSN: 2717-9494
Research Article/Araştırma Makalesi
Doi: https://doi.org/

Machine Learning Application for Diet Maintenance Period

Berrin ATALAY[1*], Sude Nur Ocak[2], Zeynep Gölbaşı[2], Elif Ezgi Yılmaz[2], Elifnaz Fidan[2]
[1]Erciyes University, Industrial Engineering Department, Assistant Prof.; https://orcid.org/0000-0002-7852-2082
[2]Erciyes University, Industrial Engineering Department, Student
* Corresponding Author: berrinucarkus@erciyes.edu.tr

Received: 24 January 2025; Accepted: 29 July 2025

Reference/Atıf: B. Atalay, "Machine Learning Application for Diet Maintenance Period", Researcher, c. 05, sy. 01, ss. 26–37.

Abstract

Objective: This study addresses one of the challenges faced during the post-diet maintenance phase: tracking nutrient intake. It aims to facilitate weight management for individuals in this period.
Materials and Methods: First, a consultation with a registered dietitian was conducted to understand the process, and sample maintenance phase meal plans were obtained. Based on this, a Windows Forms application was developed to enable individuals to track their food intake during their maintenance phase. The application helps users maintain a healthy and balanced diet during the post-diet maintenance period by considering their dietary preferences and needs and simplifying nutrient tracking. The application interface includes a user registration screen and an exchange calculation screen. Based on the data obtained from the user registration screen, the total exchange amounts can be viewed on the exchange calculation screen. By entering their daily food intake, users can see their remaining exchange rights, ensuring effective tracking during the maintenance phase.
Findings: A regression model was developed using the data collected from the application to determine whether there is a relationship between eating habits and weight. The findings indicate significant relationships between the consumption of certain nutrients and weight.
Conclusion: Based on these relationships, a valid regression model was created that can be used for future weight predictions. Additionally, the developed application allows users to track their nutrient intake during the post-diet maintenance phase, enabling them to maintain a healthy and balanced diet.

Keywords: C#, diet, maintenance period, machine learning, python

1. Introduction

Adequate and balanced nutrition is necessary to maintain a healthy life. Weight control is one of the biggest problems of our age; overweight is harmful and underweight can lead to health problems [1]. To avoid these health risks, people follow various diet programs and try to reach their goals. However, maintaining weight after weight loss is also a great challenge. In this process, nutrient monitoring is critical for individuals to lead a healthy life.

This problem encountered during the weight maintenance period is widely observed in the fields of nutrition and dietetics, health and sports. Weight control and maintenance are vital for the prevention of obesity and other health problems. Our motivation in this study is to help individuals in the weight maintenance period to lead a healthy life by providing more effective nutrient monitoring and management.

In this study, a Windows form application was developed to facilitate nutrient tracking during the post-diet maintenance period. First, the information and data sets needed in this field were obtained by interviewing an expert dietitian. This data formed the basis of the application and a Windows form application was developed using C# language. Thanks to the application, users will be able to enter the foods they eat during the day into the system to track whether they comply with the protection period program they receive under the follow-up of the dietician, and they will be able to get information about how much more they should eat from which food groups for that day.

Regression analysis was among the methods used in our study. This method was used to determine the relationship between dietary habits and weight. In particular, the effect of meat, milk, fat, bread, fruit and vegetable consumption on weight was examined and a regression model was created accordingly. The accuracy of the model was assessed by low mean square error (MSE) and high R^2 score on the test data.

In this study, the application developed for individuals in the weight maintenance period aims to help them maintain a healthy life by facilitating nutrient tracking. With the data obtained from the application, the link between dietary habits and weight was analyzed and a valid regression model was created for future weight predictions.

In the methodology section, a summary of the literature on diet, weight control, nutrient tracking, C# programming machine learning and regression modeling is presented. The problem of nutrient tracking in the post-diet maintenance period is defined and the C# application interface developed for this problem is presented. The design and development processes of the Windows form application we developed using C# programming language are explained in detail, and the process of collecting and analyzing user data is detailed. All findings are summarized and the contribution of the application to the users' healthy life is evaluated.

2. MATERIALS AND METHODS

2.1. Literature Summary

Tokgöz, Ertem, Çelik, Gökçe, Saka and Hatunoğlu investigated the eating habits of Dicle University students and investigated what percentage of them ate healthy and without skipping meals and what percentage of them ate unhealthy and skipped meals. He divided the students into groups by looking at their body mass index and examined the reasons for the difference between eating habits. They found that the main reasons for skipping meals were educational status, economic structure, income and expenditures and cultural structure. They made suggestions to improve this problem [2].

In his study, Erge, S. stated that only diet and exercise would not be sufficient in the treatment of obesity. He mentioned that in the treatment of obesity, it is essential to create a nutrition program suitable for the individual's personal characteristics, abilities and lifestyle, to balance energy intake and expenditure, and that this program should be designed in a way that the individual can feel happy and comfortable and should be supported by exercise and behavior change methods. He argued that only in this way can obese individuals achieve the desired weight loss and maintain their current healthy weight and be protected from the health problems caused by obesity [3].

In her study, Bozkurt focused on a web-based nutrition education and personal nutrition management tool designed to address nutrition problems in Turkey. This tool aims to positively change individuals' dietary habits by increasing nutritional knowledge and support personal nutrition management. The study used C#, an object-oriented programming language developed on the .NET platform. It was also stated that the study contributed to studies on personal health records by increasing knowledge and awareness on personal health management [4].

In this study, Dilmen and Ertam described the development of a computer program called Nutrition and Health Tracking (NHT), which uses Body Mass Index (BMI) to diagnose and monitor nutritional disorders in high school students. BST aims to identify and address weight and height problems in young people through early detection and interventions. BST identifies nutritional disorders through BMI calculations. The program provides diet and exercise recommendations according to BMI categories and intervenes by monitoring the student's progress. Researchers have noted that BMI does not take into account factors such as bone density and may be inaccurate in young people. Regional and cultural differences may also affect BMI interpretations [5].

In their research paper, Saylı and Akbulut discussed the development of a dietary control system to improve communication between dietitians, users and administrators. They stated that this system was

developed using SQL-based C#. They aimed to provide users with easy access to measurement and diet information and to provide access to the information they want from anywhere. They also stated that the system aims to save time and money for dieters. In general, they emphasized the importance of improving the diet management process and diet control system [6].

In their study, Dandıl and Bilen focus on the importance of work follow-up and management in universities. Their study is based on the fact that all departments can increase time efficiency in defining, monitoring and solving business processes by using a common database. Their work was carried out on Windows Azure infrastructure using C# and Asp.net programming languages. The research results emphasize the importance of using cloud computing technology for work tracking in universities to increase productivity. This study deals with a cloud computing application that can contribute to a more organized structure of work tracking processes in universities and increase productivity [7].

An important part of Aktaş and Akçay's dissertation includes an automation system software that any hospital service may need and that can meet the needs of the hospital. However, considering that the requirements may vary from hospital to hospital in SAP project studies, they developed a software product for general common requirements in this project. Visual Studio, which is used as the compiler for the C# programming language, has been examined and detailed explanations have been provided for first-time users to familiarize them with the interface components. In particular, they stated that C# is widely used in applications developed using Microsoft Visual Studio [8].

In their study, Şen, Yaşayanlar and Denizhan analyzed cargo transportation data from Turkey's sea ports. They performed data analysis on cargo statistics for the last 15 years obtained from the Ministry of Transportation and Infrastructure of the Republic of Turkey. In this analysis, they applied linear regression and artificial neural network methods using WEKA software and Python programming language and compared the results. When the accuracy rates of the forecasts were compared, they observed that the artificial neural network method gave more accurate results. The study is the first application for freight transportation data in ports and reveals that countries can increase their competitive advantage with the correct use of these data [9].

In their study, Rong and Bao-Wen used a regression model in machine learning to analyze the effect of temperature change on the sales of iced products. They built a simple linear regression model using Python programming language, taking temperature forecasts as the independent variable and sales of iced products as the dependent variable. First, they collected the previous year's temperature forecast data and the sales data of iced products, and then performed data compilation and cleaning on these data. Based on the cleaned data, they built a linear regression model to examine the relationship between independent and dependent variables. By applying the model, they were able to predict product sales based on temperature changes and with these predictions, they helped companies to adjust their production plans more flexibly. The study offers significant commercial value in determining the right production and sales strategies and provides an important theoretical basis for other companies producing iced products [10].

Yüce and Muz [12] investigated the effects of the pandemic on the dietary behavior of individuals through a questionnaire survey. They interpreted that the results were significant.

2.2. Application Interface (C# Application)

The developed C# Windows form application is a tool that will enable users to track their compliance with the nutrition plan they have determined with their dietician. The application interface is designed in such a way that the user can easily enter data and monitor daily nutrient intake. The main components of the interface are as follows:

Home screen: a landing page for users. It is given in Figure 1.

Figure 1. Landing page for users.

Sign up screen: Allows users to log in to the application. Obtains name, surname, height and weight information from the user. It stores this information in a database and stores the person's information. The height and weight information will be used for Body Mass Index (BMI) calculation. Height information should be written in cm. It is given in Figure 2.

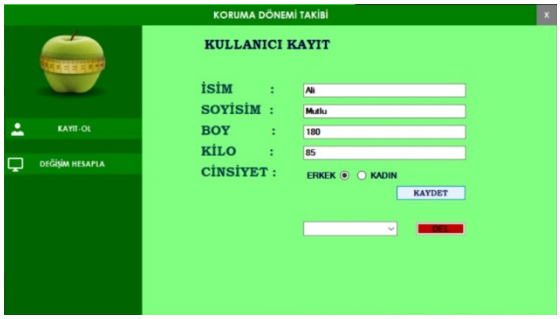

Figure 2. Sign up screen

Change calculation screen: The screen where users can enter and track their daily food consumption. On this screen, the person's information is taken and BMI is calculated. It is located between the meals and the change table. If this number is below 22, the first change table is printed, if it is above 22, the recommended daily nutrient change numbers for the person are printed using the nutrient change numbers obtained from the second change table. For each meal, there are drop-down boxes to select portions and foods. After the person has made his/her selection, he/she should add. It is given in Figure 3.

Figure 3. Application form

There is a table below the meals that gives information about the remaining change. The total change row shows the number of changes for each food group recommended for the person. The selected exchange row gives the exchange value of the nutrients when the person adds food. The remaining change row is the number of remaining nutrient changes that the person can take for that day. This information is saved in the database.

Nutrient Database: A database where users can select the information received in the registration section and the foods they eat. This database contains foods from different food groups and their nutrient exchange values. The first 100 IDs in the nutrient values are assigned to breakfast, 100-200 to lunch, 200-300 to dinner, and after 300 to snacks.

2.3. Linear Regression Analysis

Linear regression is a statistical technique used to model the relationship between a dependent variable (y) and one or more independent variables (x). It was first defined by Francis Galton in the 19th century and later expanded [11]. The main purpose of linear regression is to find a linear equation that best represents the relationship between data.

The parameters, decision variables and mathematical model of the linear regression model are as follows:

- Parameters:

$$\beta 0: Intercept$$

$$\beta 1, \beta 2, \ldots, \beta n: Coefficients\ of\ independent\ variables$$

$$\epsilon: Error\ term\ (residual\ error)$$

- Decision Variables:

$$y: Dependent\ variable\ -\ the\ value\ to\ be\ estimated$$

$x1, x2, \ldots, xn$: Independent variables - inputs used for estimation

- Mathematical Model:

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + \ldots + \beta nxn + \epsilon y\ =\ \beta 0 + \beta 1x1 + \beta 2x2 + \ldots + \beta nxn + \epsilon$$

The purpose of a linear regression model is to measure the effect of independent variables on the dependent variable. This model is widely used in different fields and can be applied to many real-world problems.

2.4. Application Problem

The study was developed in Python programming language using Jupyter Notebook IDE on Anaconda platform. Linear regression model is a statistical method used to mathematically model the relationship between dietary habits and weight and to evaluate the strength of this relationship. In this study, the effect of meat, milk, fat, bread, fruit and vegetable consumption on weight was examined and a regression model was created accordingly. The independent variables consist of the given nutrients (meat, milk, fat, bread, fruit and vegetables). The dependent variable is the user's weight. In the current system, post-diet weight maintenance tracking is done manually, which leads to difficulties in tracking. Inconsistencies and missing data in weight tracking reduce the effectiveness of the diet program and negatively affect users' motivation. Regression analysis will allow for more accurate tracking and predictions by modeling the relationship between dietary habits (meat, milk, fat, bread, fruit and vegetable consumption) and weight.

Different Directions and Constraints:

Different Aspects: The regression model can be personalized to individual dietary habits and can be applied individually for each individual's specific diet program.

Limitations: The accuracy of the model depends on the quality of the data used and external factors not considered in the model. It may also be limited to short-term data.

Acceptances:

Assumptions: It is assumed that individuals' dietary habits and metabolic rates will remain constant during the period the model is valid.

Data Collected:

Data Types (Parameters):

- Meat consumption
- Milk consumption
- Fat consumption
- Bread consumption
- Fruit consumption
- Vegetable consumption
- Daily weight measurements

Added Constraints:

Constraints:

- The data needs to be free of missing or erroneous data that could affect its accuracy.
- The model is only valid for a specific time period and should be recalibrated for longer-term forecasts.

Multiple Linear Regression Analysis

Linear regression analysis was used to model the effect of independent variables on the dependent variable. In this analysis, meat, milk, fat, bread, fruit and vegetable consumption were used as independent variables and total calories were set as the dependent variable.

The regression model is expressed in the following mathematical form:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \beta 3X3 + \beta 4X4 + \beta 5X5 + \beta 6X6 + \epsilon$$

Here the is:

$Y: User's\ weight$

$X1: Meat\ consumption\ (change)$

$X2: Milk\ consumption\ (change)$

$X3: Oil\ consumption\ (change)$

$X4: Bread\ consumption\ (change)$

$X5: Fruit\ consumption\ (change)$

$X6: Vegetable\ consumption\ (change)$

$\beta 0: Constant\ term$

$\beta 1, \beta 2, \beta 3, \beta 4, \beta 5, \beta 6: Coefficients\ of\ independent\ variables$

$\epsilon: Error\ term$

This model is used to estimate the effect of consumption of specific food groups on weight.

Data Collection

Data Required for C# Windows Form Application

The contents of the product groups between table 1 and table 6 are given. They are vegetable, fruit, bread, oil, meat and milk group contents respectively. The data in Table 1 - Table 6 were obtained in consultation with the expert dietician.

Table 1. Vegetable Group

Vegetable	Average Measure	Amount (g)
Okra	4 tablespoons (cooked)	75 g
Charleston Pepper	4 medium size	100 g
Tomato	1 small size	100 g
Tomato Juice	1 tea glass	100 g
Mallow	4 tablespoons (cooked)	150 g
Spinach	4 tablespoons (cooked)	150 g
Cauliflower	4 tablespoons (cooked)	100 g
Red Cabbage	1/8 medium size	100 g
Red Radish	5 small size or 1 medium size	100 g
Lettuce	10 leaves	100 g
Dry Onion	1 medium size	75 g
Cabbage	4 tablespoons (cooked)	100 g
Mushroom	2 tablespoons (cooked)	100 g
Lettuce	5-6 leaves	100 g
Parsley	1 medium bunch	100 g
Eggplant	4 tablespoons (cooked)	125 g
Chard	4 tablespoons (cooked)	150 g
Arugula	1 medium bunch	100 g
Cucumber	2 medium size	100 g
Purslane	4 tablespoons (cooked)	150 g
Green Beans	4 tablespoons (cooked)	150 g
Fresh Zucchini	4 tablespoons (cooked)	150 g
Watercress	1 medium bunch	100 g
Green Bell Pepper	2 medium size	100 g
Green Chili Pepper	10 medium size	100 g
Green Onion	3-4 medium size	75 g

Table 2. Fruit Group

Fruit	Average Measure	Amount (g)
Raspberry	1/2 cup of water	75 g
Pineapple	1 thin slice	75 g
Pear	1 medium size	100 g
Quince	1/4 medium size	80 fr
Blackberry	1/2 su bardağı	75 g
Strawberry	12 pieces	175 g
Lemon	1 medium size	100 g
Apple	1 little size	100 g
Plum	5 pieces	100 g
Grapefruit	1/2 medium size	125 g

Tangerine	1 large size	100 g
Green Plum	10 pieces	50 g
Loquat	6 pieces	125 g
Apricot	3 pieces	100 g
Red Plum	5 pieces	50 g
Cherry	12 pieces	75 g
Kiwi	1 medium size	120 g
Damson Plum	3-4 pieces	125 g
Pomegranate	1/2 little size	80 g
Orange	1 medium size	100 g
Peach	1 medium size	100 g
Bitter Orange	1 medium size	100 g
Cherry	14 pieces	80 g
Watermelon	1 thin slice	200 g
Melon	1 thin slice	200 g
Mulberry	8 pieces	50 g
Baghdad Date	1/2 medium size	50 g

Table 3. Bread Group

Bread	**Average Measure**	**Amount (g)**
Bread	1 thin slice	25 g
Lentil Soup	1 bowl	20 g
Tarhana Soup	1 bowl	20 g
Wheat Flour Soup	1 bowl	20 g
Tomato Soup	1 bowl	20 g
Rice Soup	1 bowl	15 g
Ezogelin Soup	1 bowl	15 g
Rice Pilaf	3 Tablespoons	20 g
Bulgur Pilaf	3 Tablespoons	20 g
Couscous	3 Tablespoons	20 g
Pasta	3 Tablespoons	20 g
Noodles	3 Tablespoons	20 g
Potato	1 little size	90 g
Chestnut	2 medium size	30 g
Popcorn	1 cup of water	20 g
White Chickpeas	1 tea glass	50 g
Yellow Chickpeas	1 tea glass	50 g
Etimek	2 slice	18 g
Grisin	3 pieces	20 g
Thin Cracker	13-18 pieces	18 g
Cracker	5 pieces	22 g
Breadcrumbs	2 pieces	19 g
The breadcrumbs are plain, large.	1 piece	20 g
breadcrumbs, plain, small	7 pieces	20 g

Goldenrod with bran	4,5 pieces	20 g
Sesame Stick Kıraker	4 pieces	20 g

Table 4. Oil Group

Oil	Average Measure	Amount (g)
Sunflower Oil	1 dessert spoon	5 g
Walnut	1-2 pieces	8 g
Hazelnut	5-6 pieces	10 g
Hazelnut Oil	1 dessert spoon	5 g
Corn Kernel Oil	1 dessert spoon	5 g
Soybean Oil	1 dessert spoon	5 g
Olive	5 pieces	15 g
Olive oil	1 dessert spoon	5 g
Avocado	1/4 pieces	15 g
Butter	1 dessert spoon	5 g
Margarine (soft)	1 dessert spoon	5 g
Sunflower Seeds	1 tablespoon	10 g

Table 5. Meat Group

Meat	Average Measure	Amount (g)
Fish meat (lean)	About 1 meatball	30 g
White Cheese	1 matchbox	30 g
Steak	About 1 meatball	30 g
Turkey Meat	About 1 meatball	30 g
Cheddar cheese	2/3 of a matchbox	30 g
Minced meat	About 1 meatball	30 g
Meatball	1 piece	30 g
Cubes	3-4 pieces	30 g
Curd Cheese	2 tablespoons	30 g
Cutlet (boneless)	About 1 meatball	30 g
Chicken meat (skinless)	About 1 meatball	30 g
Egg	1 piece	30 g

Table 6. Milk Group

Milk	Average Measure	Amount (g)
Ayran	2 cups	320 g
Milk	1 cup	160 g
Yoghurt	1 cup	160 g
Kefir	1 cup	200 g

The data received is a sample protection diet list. Afterwards, tables indicating the types and portions of food that users should eat after the diet were taken.

If the body mass index value is below 22, the recommended change values are as follows: 3 changes in the milk group, 6 changes in the meat group, 4 changes in the bread group, 4 changes in the vegetable group, 2 changes in the fruit group and 4 changes in the fat group.

If the body mass index value is above 22, the recommended change values are as follows: 2 changes from the milk group, 5 changes from the meat group, 3 changes from the bread group, 3 changes from the vegetable group, 1 change from the fruit group and 3 changes from the fat group.

Data Required for Linear Regression

Among a total of 10 participants, 2 people with BMI values above and below 22 were selected and daily data were collected for a total of 30 days. Observations were made in May-June 2024. Participants entered their data themselves using the application. These data were used to determine the relationships between eating habits and weight during the post-diet maintenance period and a regression model was created. The accuracy of the model was assessed by low mean square error (MSE) and high R^2 score.

One meat exchange contains 6 grams of protein, 5 grams of fat and provides 69 calories of energy. A milk exchange contains 6 grams of carbohydrates, 4 grams of protein and 3.5 grams of fat and provides 71 calories. A butter exchange contains 5 grams of fat and provides 45 calories of energy. A bread exchange contains 15 grams of carbohydrates and 2 grams of protein, providing 68 calories of energy. One fruit exchange contains 12 grams of carbohydrates and provides 48 calories of energy. A vegetable exchange contains 6 grams of carbohydrates, 1 gram of protein and provides 28 calories of energy.

The regression model was constructed using the following independent variables:

- Meat consumption (30 grams) = 1 Change of meat
- Milk consumption (160 grams) = 1 Exchange milk
- Fat consumption (5 grams) = 1 Change fat
- Bread consumption (25 grams) = 1 change of bread
- Fruit consumption (100 grams) =1 Change fruit
- Vegetable consumption (100 grams) = 1 Change vegetables

3. FINDINGS

An application was coded in C# for the protection period tracking problem and real user data was obtained from this application. The number of changes given according to the BMI (Body Mass Index) value calculated using the data obtained from the user was used. 10 different users were tested. Nutrient changes such as meat, milk, etc. given between Tables 1 and 6 constituted the parameters of the problem. User information and nutrient change values are extracted from databases created using SQL. The user registers by entering his/her information on the sign-up screen. It switches to the change calculation tab and can see how much change is left from which food group by selecting the foods and portions consumed during the day. Among a total of 10 participants, 2 people with BMI values above and below 22 were selected and daily data were collected for a total of 30 days

Two studies were conducted for linear regression with Python coding, two linear regression models obtained as a result of 30-day data for BMI values above and below 22. When the correlation matrix for the relationship between nutrients and total calories is analyzed.

For those with a BMI below 22, the nutrient that most affects the total calorie intake and weight of the person during the day is meat. Meat is followed by bread, fruit, milk, vegetables and fat. When training the model, five models with test dimension 0.1-0.2-0.25-0.3-0.4 were run for training and test data. The most significant model is model 1 with an R2 value of 0.95 at a test size of 0.1. As a result of the regression analysis, it was found that the consumption of certain food groups had significant effects on weight. In particular, meat and bread consumption had more significant effects on weight.

For BMI over 22, the total calories and weight of the person during the day. The nutrient that affects the most is the meat exchange. The meat nutrient is followed by fruit, milk, bread, butter and vegetables. Training and testing when training the model on five models with a test size of 0.1-0.2-0.25-0.3-0.4 for the data was studied. The most significant model at test size 0.1, R2 value of 0.79with model 1. As a result of the regression analysis, it was found that certain food groupsconsumption has been found to have significant effects on weight. Especially meat and fruit consumption has been observed to have more pronounced effects on weight.

4. DISCUSSION

At the beginning of the project, an application aimed at facilitating the monitoring of the maintenance period was successfully designed. Individuals entered the details of their meals into the form application, which made it easier to track their maintenance period. During the regular diet phase, individuals are provided with ready-made diet lists, while during the maintenance period, they are only given information about the quantities they should consume daily, and they are expected to balance their intake accordingly. With the developed Windows application, users were able to log the items they consumed and achieve a balanced intake. The application, which is already functioning correctly, can be further developed with the addition of a database. For example, by adding a Dietitian module, users' data could be shared with a dietitian, enabling the dietitian to make adjustments or suggest different quantities for the user's consumption in other periods.

Another development option could involve offering sample menus based on the remaining exchange counts. For instance, if an individual needs to consume 5 portions of meat, 3 of vegetables, 2 of fruit, 2 of bread, 2 of fat, and 3 of dairy daily, and they consumed 1 portion of meat, 1 of vegetables, 1 of fruit, 1 of bread, and 1 of dairy at breakfast, the application could use the logged data to present alternative sample menus based on the remaining exchanges.

In the Windows form application, data for 30 days (from the same individual) was retrieved from the database, and multiple regression analysis was applied based on the individual's consumption. Regression analysis was performed for two users with different BMI values. This analysis supported the study by helping users better understand the relationship between their eating habits and weight changes. For users with a BMI value below 22, the R^2 value obtained from Model 1 with a 0.1 test size was 0.95, making it the most meaningful model explaining the relationship between nutritional components and total calorie intake. For users with a BMI value above 22, the R^2 value from Model 1 with a 0.1 test size was 0.79. Since these results are close to 1, it shows that both models are highly successful in estimating total calorie intake.

The effectiveness of the application was evaluated through the regression model, enabling users to monitor their adherence to the maintenance period program prescribed by their dietitian.

5. CONCLUSION

A new application has been developed to help individuals in the diet maintenance phase monitor the progress of their diets. Users can log the details of what they eat in a meal into the application, and it will instantly calculate the remaining quantities they can consume for subsequent meals. Since the data is stored in a database, a dietitian module can be added, allowing this information to be shared with a dietitian.

With the data stored in the application's database, the user's nutrition program can be adjusted on a weekly or daily basis by adding or reducing certain items.

In this study, data stored in the database for individuals with BMI values above and below 22 were analyzed. Using these data, a multiple linear regression model, a machine learning technique, was built to examine the impact of food groups (meat, dairy, fat, fruit, vegetables, and bread) on calorie intake. The results obtained optimized the individual's total calorie intake. According to the regression results, the R^2 value was 0.95 for the recommended model for individuals with a BMI value below 22, while it was 0.79 for those with a BMI above 22. As a result, the proposed regression model yielded statistically significant outcomes.

Acknowledgments

The study was supported within the scope of TÜBİTAK 2209-A University Students Research Projects Support Program 2023. We would also like to thank Dietitian Sıdıka Merve Okur for her contribution to the study.

Contribution of Researchers

Study idea: BA, Study design: SO, ZG, EF and EEY; Obtaining data: SO, ZG, EF and EEY; Analyzing data: BA, SO, ZG, EF and EEY; Creating the article draft BA; Critical review for content: BA, Final approval of the version to be published: BA.

Conflicts of Interest

Authors declared no conflict of interest.

Ethics committee approval

Ethics committee approval is not required for this study.

References

[1] https://hsgm.saglik.gov.tr/tr/beslenme/yeterli-ve-dengeli-beslenme.html

[2] Tokgöz, P., Ertem, M., Çelik, F., Gökçe, Ş., Saka, G., & Hatunoğlu, R. (1995). Üniversite öğrencilerinin beslenme alışkanlıklarının saptanmasına ilişkin bir araştırma. Beslenme ve Diyet Dergisi, 24(2), 229-238.

[3] Erge, S., (2003). Obezlerde yeme davranışı ve obezitede davranış tedavisi İlkeleri. Beslenme ve Diyet dergisi, 32(1), 47-59.

[4] Bozkurt, S. (2008). Beslenme ve diyet web portalı.

[5] Dilmen, H., & Ertam, F. (2013). Online tracking nutrition and health of high school students. The Online Journal of Science and Technology, 3(1), 119-123.

[6] Saylı, A., Akbulut, C., (2017). Analysis, design and implementation of diet control system. Avrupa Bilim ve Teknoloji Dergisi, 7(11), 2-6.

[7] Dandıl, E., & Bilen, M. S. (2013). Üniversitelerde iş takibi için bir bulut bilişim uygulaması. Akademik Bilişim Konferansı, 23-25.

[8] Aktaş, F. & Akçay, M., (2013). Hastane Otomasyon Projesi. T.C. Dumlupınar Üniversitesi, Bilgisayar Mühendisliği Bölümü Lisans Bitirme Tezi, Kütahya, 46.

[9] Şen, H., Yaşayanlar, A., & Denizhan, B. (2021). Türkiye'de Limanlarda Yük Taşımacılığı Veri Analizi. Avrupa Bilim ve Teknoloji Dergisi, (22), 436-444.

[10] Rong, S., & Bao-Wen, Z. (2018). The research of regression model in machine learning field. In MATEC Web of Conferences (Vol. 176, p. 01033). EDP Sciences.

[11] Galton, F. (1886). Regression towards mediocrity in hereditary stature. The Journal of the Anthropological Institute of Great Britain and Ireland, 15, 246-263.

[12] Yüce, G. E., & Muz, G. (2021). COVID-19 pandemisinin yetişkinlerin diyet davranışları, fiziksel aktivite ve stres düzeyleri üzerine etkisi. Cukurova Medical Journal, 46(1), 283-291.

ANKARA SCIENCE UNIVERSITY, RESEARCHER
Vol. 5, No. 1, July 2025
e-ISSN: 2717-9494
Research Article/Araştırma Makalesi
Doi: https://doi.org/

Bridging the Language Gap in RAG: A Case Study on Turkish Retrieval and Generation

Erdoğan BIKMAZ[1*], Mohammed BRIMAN[2], Serdar ARSLAN[3]
[1]Çankaya University, Department of Computer Engineering, Ankara, Türkiye; https://orcid.org/0009-0006-7147-4539
[2]SmartICT Bilişim A.Ş., Ankara, Türkiye; https://orcid.org/0009-0000-5785-6916
[3]Çankaya University, Department of Computer Engineering, Ankara, Türkiye; https://orcid.org/0000-0003-3115-0741

* Corresponding Author: ebikmaz@smartict.com.tr

Received: 14 May 2025; Accepted: 22 May 2025

Reference/Atıf: E. Bikmaz, M. Briman, ve S. Arslan, "Bridging the Language Gap in RAG: A Case Study on Turkish Retrieval and Generation", Researcher, c. 05, sy. 01, ss. 38–49.

Abstract

With the rise of Large Language Models (LLMs) and LLM-based Retrieval-Augmented Generation (RAG) systems, there is a high demand for developing RAG applications that utilize LLM reasoning capabilities for handling intensive text systems in multilingual settings. However, RAG components are primarily developed for the English language, which hinders their ability to retrieve and construct precise multilingual information for LLMs to answer, especially for the Turkish language. In this work, we aim to explore the effects of developing comprehensive RAG systems that handle Turkish question-answer retrieval and generation tasks. We experiment with fine-tuning two major components on Turkish data: the embedding model used for data ingestion and retrieval, and a reranker model that ranks the retrieved documents based on their relevance to a query. We evaluate four RAG systems using six evaluation metrics. Experimental results show that fine-tuning retrieval components on Turkish data improves the accuracy of LLM responses and leads to improved context construction.

Keywords: retrieval augmented generation, large language models, embedding, information retrieval

1. Introduction

Large Language Models (LLMs) have revolutionized the landscape of Natural Language Processing (NLP) due to their ability to perform complex tasks simply by being instructed through prompts. The Transformer architecture [1] enabled the emergence of capable LLMs such as GPT-4 [2], Claude 3.7 Sonnet [3], LLaMA models [4], Mistral [5], and Phi [6]. Although LLMs have demonstrated powerful reasoning capabilities through various prompting techniques [7, 8], they still suffer from hallucination problems [9]—that is, they may generate false information if the input prompt requests content that does not exist in the LLMs' training data. Thus, the hallucination problem raises concerns about the reliability of LLMs when deployed in real-world applications.

Several approaches have been proposed to reduce LLM hallucinations, with one of the most prominent and widely adopted being Retrieval Augmented Generation (RAG) [10]. RAG aims to mitigate hallucinations by incorporating external information into the prompt as "context." This added context helps the LLM provide answers involving out-of-domain knowledge. A typical RAG system consists of two main components: a retriever, which retrieves relevant embeddings [11] of context documents based on the embedding of a user query; and a generator, which uses the retrieved context along with the query to generate a response. The first stage in developing a RAG system is data ingestion, where large raw texts are split into smaller chunks, embedding vectors [11] are generated for each chunk, and these vectors are stored in a Vector Database (VDB) [12] for later semantic retrieval [13]. Semantic search refers to a similarity-based retrieval operation between the embedding vector of a user query and the document embeddings stored in the VDB. Therefore, the quality of the embedding vectors and the degree of contextual preservation between text chunk embeddings significantly influence the overall performance of a RAG system.

One of the biggest obstacles in deploying RAG applications is the limited multilingual representation for languages other than English [14]. There is a growing demand for developing RAG systems tailored to multilingual settings, particularly for the Turkish language. Although significant progress has been made in multilingual LLMs [15-16], there remains a clear lack of high-quality Turkish language support in retrieval components and embedding models. Furthermore, another critical limitation lies in the development of retrieval components that are both multilingual and capable of handling multiple domains [17].

In this work, we aim to improve the performance of RAG systems on Turkish data by fine-tuning two retrieval components: the embedding model and the reranker. Additionally, we experiment with a prepositional chunking method [18] to enhance the quality and contextual relevance of the retrieved chunks.

Our contributions can be summarized as follows:

- We fine-tune two retrieval components—a sentence embedding model and a cross-encoder reranker—on Turkish data.
- We develop four distinct RAG systems to evaluate their performance with the fine-tuned retrieval components.
- We compute six different metrics for each of the four RAG systems to assess various aspects of their performance.

2. Related Work

Examples of multilingual and domain adaptation in LLM-based RAG applications include the evaluation of RAG systems for health-related chatbots in Indian languages, which involved analyzing the performance of several multilingual LLMs [19]. However, that experiment employed a standard embedding model from OpenAI. Xu et al. [20] adapted RAG to a Chinese medical analysis task using a two-stage retrieval process and a specialized Chinese text segmentation method. For semantic search, they utilized a QWEN-based LLM [16].

Retrieval Augmented Fine-Tuning (RAFT) [21] combines supervised fine-tuning with RAG to adapt systems to domain-specific knowledge. Rameel et al. [22] investigated RAG improvements in multilingual settings for real-world applications. Specifically, they compared paragraph-based, semantic-unit-based, topic modeling, and entity-based text splitting methods. Their findings highlighted the importance of balancing chunk size and overlap to preserve relevant information.

There are multiple methods for training and adapting embedding models to support multilingual capabilities. Early word embedding approaches explored various algorithms, such as adversarial training and pseudo-supervised refinement, to improve multilingual word-level embeddings [23]. However, the dominant approach for RAG systems is the use of full-sentence embeddings generated by Transformer-based pre-trained encoder models. Notable works in sentence embeddings include the early multilingual Universal Sentence Encoder developed by Yang et al. (2019) [24], which included Turkish in its training dataset, as well as multilingual SBERT models [11].

There are several multilingual embedding models categorized by parameter count, embedding dimensionality, and performance on the MTEB leaderboard [25-26]. Among the leading models is MiniLM [27], which uses a knowledge distillation teacher–student approach [28], applying distillation on the last layer of the Transformer teacher model to learn embedding representations. MiniLM includes multilingual data from a wide range of languages. The General Text Embedding (GTE) family of models [29] employs multi-stage contrastive learning across a diverse set of datasets. GTE models are based on a 110M-parameter BERT backbone [30], and several variants support up to 70 languages. Specifically for retrieval and RAG use cases, Wang et al. (2022) [31] introduced the e5 family of embedding models, which are tailored for RAG-style retrieval. These models use special input prefixes—“query:” for user queries and “passage:” for context documents. Additionally, e5 models are trained on multilingual datasets covering approximately 27 languages.

Reranker models are typically based on BERT-style cross-encoder architectures [30], which capture semantic relationships between queries and documents. One example is the Standalone Neural Reranking Model (SNRM), developed by Zamani et al. [32], which uses high-dimensional sparse representations for query–document pairs to perform retrieval. ColBERT, proposed by Khattab et al. [33], introduced a BERT-based reranker that employs contextualized late interaction between queries and documents to improve reranking efficiency. For fine-tuning reranker models for specific tasks, Moreira et al. [34] from NVIDIA explored the fine-tuning of both cross-encoders and decoder-based rerankers. Their results showed that fine-tuning decoder-based reranker models leads to improved accuracy compared to baseline rerankers. The development of Turkish BERT models by Kesgin et al. [35] enabled further exploration of multilingual adaptations. A notable example is the cross-encoder Turkish reranker model "turkish-colbert" [36], which was fine-tuned from a Turkish BERT cross-encoder using a Turkish-translated version of the MS MARCO dataset [37].

3. Methodology

We aim to improve the performance of RAG systems in the Turkish language by focusing on maximizing the relevance of the contextual information generated by the RAG pipeline. To achieve this, we experiment with three key components of the RAG architecture: (1) fine-tuning an embedding model on Turkish sentence pairs, (2) fine-tuning a reranker model on Turkish query–document pairs, and (3) applying a prepositional chunking strategy during the data ingestion stage to further enrich document context. We develop four distinct RAG systems, each incorporating a different combination of these components.

3.1 Fine-tuning an Embedding Model

We fine-tuned an embedding model to adapt a pre-trained multilingual model for encoding Turkish language data, enabling it to effectively embed novel and domain-specific sentences. We used an open Turkish RAG dataset containing 6,000 source documents and corresponding QA pairs for each document [38]. The significance of the QA pairs lies in the availability of ground-truth answers for each question related to the source documents.

Our embedding model of choice was multilingual-e5-large [31], selected for its multilingual capabilities, compact 384-dimensional embeddings, and suitability for retrieval tasks. The fine-tuning process involved several data preprocessing stages, including text cleaning, normalization, punctuation removal, and structuring the dataset into the required ["question," "context"] pair format for a RAG system. Additionally, model-specific prefixes were added to distinguish queries from contexts. We employed Multiple Negative Ranking as the loss function and Adam as our optimizer. Training parameters are summarized in Table 1. Finally, we will refer to our model as multilingual-e5-tr-rag.

Table 1. Fine-tuning hyperparameters for the embedding model

Parameter	Value
Base Model	multilingual-e5-large
Embedding dim	384
Training data size	5399
Learning rate	2e-5
Loss function	Multiple Negatives Ranking Loss
Evaluation metric	Recall@k (k=10)

3.2 Fine-tuning a Reranker Model

We fine-tuned a cross-encoder reranker model on a Turkish RAG dataset to improve the accuracy of reranking Turkish documents. For this purpose, we selected the "jina-reranker-v2-base-multilingual"

reranker model developed by [39] as our base model due to its multilingual capabilities. Additionally, we fine-tuned the model using resources provided by the sentence-transformer library [11].

To prepare the reranking dataset, we used the queries and documents from the original dataset [38] and applied the "Hard Negative Mining" (HNM) process [40], which selects documents that may appear to be relevant to a given query but are not. This process improves the reranker model's ability to differentiate between relevant and irrelevant query-document pairs. The resulting dataset consists of (query, document, label) triples, with a binary label column indicating whether the query-document pairs are relevant.

To evaluate the fine-tuning results, we utilized the BEIR retrieval evaluation benchmark [41]. Our fine-tuned model will be referred to as jina-reranker-multilingual-wiki-tr-rag.

Table 2. Fine-tuning hyperparameters for the reranking model

Parameter	Value
Base Model	jina-reranker-v2-base-multilingual
Embedding dim	1024
HNM training data size	26004
Training data size	5399
Learning rate	2e-5
Loss function	Binary Cross Entropy Loss
Evaluation metric	Cross-Encoder Nano BEIR Evaluator

3.3 Evaluation Metrics

This experiment aims to evaluate the quality of RAG systems in terms of generating factual answers and measuring the utilization of retrieved context. We assessed each of the four RAG systems from three perspectives: the relationship between the generated LLM answer and the retrieved contexts, the relevance between the queries and the retrieved contexts, and the similarity between the LLM-generated answer and the ground truth answer. We used RAG-specific metrics from the Retrieval Augmented Generation Assessment (RAGAS) [42] and standard NLP metrics such as ROUGE-N [43] and BERTScore [44].

To assess the similarity between the LLM-generated answers and the ground truth, we employed ROUGE-N and BERTScore. Both ROUGE-N and BERTScore compute recall, precision, and F1 scores between two texts (e.g., LLM-generated answer and ground truth). ROUGE-N calculates the similarity based on overlapping n-gram units, while BERTScore uses BERT-based embeddings to evaluate similarity, considering semantic meaning.

For RAG-specific evaluations, we utilized four RAGAS metrics: Faithfulness, Answer Relevance, Context Recall, and Context Precision.

- **Faithfulness**: Faithfulness measures the factual consistency between the LLM-generated answer and the retrieved context.
- **Answer Relevance**: This metric evaluates the relevance between the LLM-generated answer and the original question.
- **Context Recall**: Context Recall assesses whether the necessary documents have been retrieved to answer the question. It is computed by comparing the retrieved context to the ground truth.
- **Context Precision**: Context Precision measures the signal-to-noise ratio between the question and the retrieved context.

3.3 RAG Systems

We developed a RAG system [10] consisting of two main stages: the data ingestion stage, where data undergoes a splitting process to break down large texts into smaller, manageable chunks, and the retrieval and generation stage, where a user query triggers a retrieval process to fetch the most relevant text chunks. All information is placed into a prompt template containing both the original user query and the retrieved chunks (see the prompt template in Table A1). The final prompt is then passed to an LLM for generating the answer. The RAG setup is illustrated in Figure 1.

In the data ingestion stage, we used both Recursive and Prepositional chunking methods in our RAG systems to compare their performance. We compared the effectiveness of a simple recursive separator-based splitting method with an LLM-based chunking method. Our goal was to explore how much context is preserved between the two methods and how the resulting text chunks impact the quality of the retrieval stage.

Recursive chunking [45] is a standard text-splitting method that uses separators in its process. It splits the text recursively until the smallest possible sentence is reached. However, due to its simplicity, some chunks may lose important context.

Prepositional chunking [18] is a model-based chunking method in which an LLM is prompted to remove ambiguity from a piece of text by following a set of chunking instructions. As the name suggests, prepositional chunking creates factual, concise, and self-contained sentences, with the aim of improving retrieval results and helping LLMs generate more accurate responses. An example is given in Table 3.

Table 3. Prepositional chunking applied to a piece of Turkish text

Original Chunk	Prepositional Chunks
San Miguel de Allende, Meksika'nın iç bölgelerinde yer alan Guanajuato Eyaleti'nin doğu kesiminde bulunan bir şehir ve belediyedir. Bajío'nun makrobölgelerinin bir parçasıdır. Meksiko'dan 274 km (170 mi), Guanajuato Eyaleti'nin başkentinden 97 km (60 mi) uzaklıkta yer alır. Eski adı San Miguel el Grande olan şehir, 1826'da yapılan değişikle Ignacio Allende'nin anısına günümüzdeki adı almıştır.	San Miguel de Allende, Meksika'nın iç bölgelerinde yer almaktadır. San Miguel de Allende, bir şehir ve belediyedir. San Miguel de Allende, Bajío'nun makrobölgelerinin bir parçasıdır.

We applied prepositional chunking to our documents using "gemini-2.0-flash-lite" [46] as our LLM of choice due to its multilingual capabilities. Table 3 provides an example of prepositional chunking applied to a Turkish text. The prompt used to apply the chunking process is provided in Table A1 in Appendix A.

To experiment with the effects of embedding and reranker models on the performance of a RAG system, we developed four different versions, each employing a combination of text splitting and embedding models. We use two different text splitting methods: a simple recursive method with a defined set of separators, and a prepositional chunking method that utilizes an LLM to split the text into self-contained, factual sentences. Additionally, we use "gemma-3-27b-it" [47] as the primary generation LLM due to its multilingual capabilities.

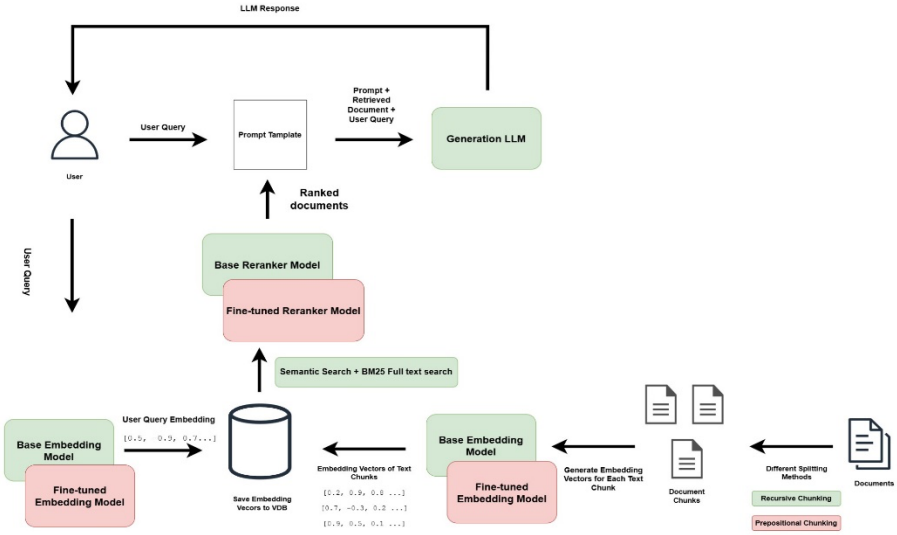

Figure 1. RAG system overview. The green boxes show the base version of each component and the pink boxes show our custom componants.

We used a hybrid search pipeline for the retrieval process, combining a cosine similarity-based semantic search retriever with a BM25 keyword retriever [48]. All vector embeddings were stored in a Chroma DB vector database [49].

The configurations of the RAG systems are as follows:

- **Base RAG:** Recursive chunking + paraphrase-multilingual-MiniLM-L12-v2 base embedding model [50] + semantic search + ColBERTv2.0 reranking [33].
- **RAG V1:** Recursive chunking + fine-tuned multilingual-e5-tr-rag embedding model [50] + hybrid search + ColBERTv2.0 reranking [33].
- **RAG V2:** Recursive chunking + fine-tuned multilingual-e5-tr-rag embedding model [50] + hybrid search + fine-tuned jina-reranker-multilingual-wiki-tr-rag reranking.
- **RAG V3:** Prepositional chunking + fine-tuned multilingual-e5-tr-rag embedding model [50] + hybrid search + fine-tuned jina-reranker-multilingual-wiki-tr-rag reranking.

The RAG evaluation dataset consists of 600 samples taken from the test set of the Turkish RAG dataset [38]. Each sample includes three columns: Context, Question, and Answer. The Context represents the source document, the Question is a query related to the Context, and the Answer is the corresponding ground-truth response.

4. Experimentation Setup

In terms of datasets, we use 600 test samples from the Turkish RAG dataset [38]. For evaluation purposes, the dataset was transformed into an evaluation format consisting of the following columns:

- **Question:** Questions about the information presented in the source documents.
- **Answer:** The LLM-generated answer from the RAG system.
- **Contexts:** The set of retrieved documents via Semantic Search.
- **Ground Truth:** The ground truth answers present in the dataset.

Each of the four metrics presented in Section 3.3—namely, Faithfulness, Answer Relevance, Context Recall, Context Precision, ROUGE-N, and BERTScore—is applied to the outputs of each of the four RAG systems.

4. Experimentation Results and Analysis

By examining the results in Figure 2, we observe that RAG V2 outperforms the other three configurations, including the baseline RAG system. Notably, RAG V2 also produces higher-quality LLM-generated responses, as indicated by the ROUGE-N and BERTScore metrics. These findings suggest that fine-tuning retrieval components—such as embedding models and rerankers—on domain-specific language significantly enhances the retrieval effectiveness in RAG systems. Consequently, the improved relevance and quality of the retrieved context lead to more accurate and informative responses generated by the LLM.

Interestingly, Figure 2 also reveals that RAG V3, which incorporates the prepositional chunking method, performs worse than both the baseline and the other RAG configurations. We hypothesize that the use of short, self-contained sentences—while intended to enhance clarity—may negatively impact the richness of the contextual information captured by the embedding vectors. As a result, this could lead to suboptimal retrieval, with relevant documents being missed due to insufficient contextual cues.

An examination of the RAGAS Faithfulness scores [42] for all four systems, as presented in Figure 2, reveals that RAG V3 exhibits the lowest level of factual consistency between the LLM-generated responses and the retrieved context. This metric, which measures the degree to which the generated answer aligns with the retrieved supporting evidence, indicates that the prepositional chunking method used in RAG V3 may hinder the model's ability to ground its responses in contextually accurate information compared to the other configurations.

$$Faithfulness\ Score = \frac{\text{Number of claims in the response supported by the retrieved context}}{Total\ number\ of\ claims\ in\ the\ response} \quad (1)$$

Tables A2 and A3 in Appendix A illustrate the outputs of the RAG V2 and RAG V3 systems for a representative sample from our dataset [38]. In Table A2, RAG V2 achieves a Faithfulness score of 1.0, with the response explicitly grounded in the first retrieved chunk—denoted by "(1)" —demonstrating effective alignment between context and answer. The recursive chunking method employed by RAG V2 produced semantically rich and coherent chunks, enabling the LLM to generate a well-supported response. In contrast, Table A3 shows that RAG V3, which utilized prepositional chunking, retrieved short, atomic sentences that often lacked sufficient contextual information. Although prepositional chunking aims to produce concise and self-contained units, in this case it resulted in lower-quality context that negatively affected the LLM's output. This outcome highlights that, despite the use of fine-tuned retrieval components, the quality of the initial chunking strategy can significantly influence the overall performance of a RAG system.

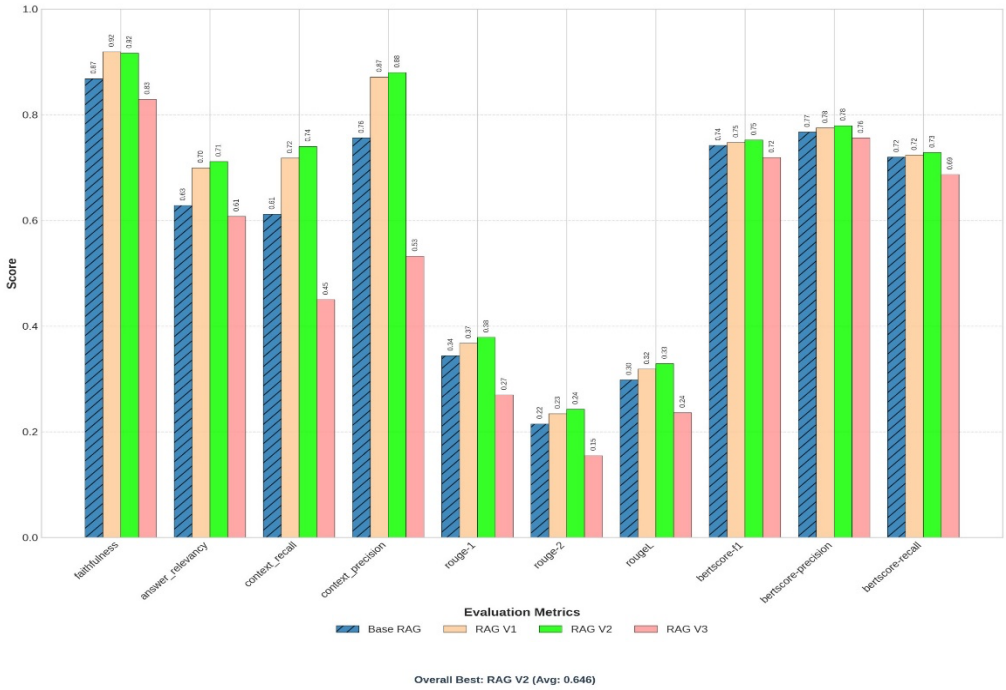

Figure 2. RAGAS, ROUGE-N, and BERTScore metric results for each RAG system

6. Conclusion

In this work, we explored the impact of fine-tuning retrieval components for a domain-specific language, such as Turkish, on improving both retrieval performance and LLM-generated response quality. We fine-tuned a multilingual embedding model and a reranker model on Turkish text with the goal of enhancing both embedding quality and reranking effectiveness. Additionally, we examined the effectiveness of a novel chunking method called prepositional chunking. To evaluate the impact of these components, we developed four different RAG systems and assessed their performance using six key metrics: faithfulness, answer relevance, context recall, context precision, ROUGE-n, and BERTScore. Our experimental results demonstrate that further adapting multilingual base models to domain-specific languages significantly improves retrieval performance on the target tasks. We show that further fine-tuning multilingual embedding and reranking models drastically aligns the embedding space to our specific Turkish language domain and enhances the retrieval quality for better LLM generation. Moreover, we demonstrate that using a standard chunking method with custom fine-tuned retrieval components performs better than using LLM-based chunking methods such as prepositional chunking. Future studies will focus on combining both embedding (bi-encoders) and reranking (cross-encoder) architectures to create fusion retrieval and reranking models for improving the accuracy of retrieval components for the Turkish Language.

Appendix A: Tables and Figures.

Table A1. Prepositional chunking prompt in Turkish

Prepositional Chunking Prompt
Lütfen aşağıdaki metni basit, kendi içinde anlamlı önermelere ayırın. Her bir önerme aşağıdaki kriterleri karşılamalıdır:

1. Tek Bir Gerçeği İfade Etmeli: Her önerme, belirli bir gerçeği veya iddiayı ifade etmelidir.
2. Bağlama İhtiyaç Duymadan Anlaşılır Olmalı: Önerme, başka bir bağlama ihtiyaç duymadan anlaşılır olmalıdır.
3. Zamir Değil, Tam İsimler Kullanılmalı: Zamirler veya belirsiz ifadeler yerine tam varlık isimleri kullanılmalıdır.
4. İlgili Tarih/Sıfatları İçermeli: Uygunsa, gerekli tarih, zaman ve niteleyiciler eklenerek önerme kesin hale getirilmelidir.
5. Tek Özne-Yüklem İlişkisi İçermeli: Tek bir özne ve ona karşılık gelen bir eylem veya özellik üzerine odaklanılmalı; bağlaçlar veya birden fazla yargı içermemelidir.

Table A2. Faithfulness Evaluation Example for RAG V2 sample

Parameter	Value
User Input	Havayolu Serbestleştirme Kanunu, havacılık endüstrisinde hangi değişiklikleri beraberinde getirdi?
Response	Havayolu Serbestleştirme Kanunu, bilet fiyatları, rotalar ve yeni havayollarının pazara girişi üzerindeki devlet kontrolünü kaldırdı ve yolcuları piyasa güçlerine maruz bıraktı (1).
Retrieved Contexts	Airline Deregulation Act (Türkçe: Havayolu Serbestleştirme Kanunu), 1978'de Amerika Birleşik Devletleri başkanı Jimmy Carter tarafından imzalanan, ticari havacılık'daki bilet fiyatları, yolları ve yeni havayolların pazara girişi üzerindeki devlet kontrolünü kaldırma amaçlı federal kanun. Civil Aeronautics Board'un denetleme gücü yavaş yavaş azaltılmış, sonunda havacılık endüstrisinde yolcuları piyasa güçlerine maruz bırakmıştır. Fakat kanun, Federal Aviation Administration'ın (FAA) tüm hava emniyet yönleri. Pseudostaurastrum, ışınlılar grubunda yer alan bir Protista. Aralarında 3-4 boşluk olan her bir pseudostaurastrum'un boşlukları birbirine eşittir. Enternasyonal turuncu havacılık endüstrisinde nesneleri çevresinden ayırt etmek için kullanılan, güvenlik turuncusuna benzeyen, ama daha derin ve kırmızı bir tona sahip olan renktir. Liberizm ya da özgürcülük, filozof Benedetto Croce tarafından kullanılan ve İtalyan-Amerikan siyaset bilimci Giovanni Sartori tarafından İngilizce olarak popüler olan laissez-faire kapitalizminin ekonomik doktrini için kullanılan bir terimdir. Ekonomik doktrin laissez-faire kapitalizmini ima eden terim, filozof Benedetto Croce tarafından ilk kez kullanılmıştır ve siyaset bilimci Giovanni Sartori tarafından popülarize edilmiştir. Kavram İtalyanca liberismo kelimesinden türetilmiştir.
Faithfulness	1.0

Table A3. Faithfulness Evaluation Example for RAG V3 sample

Parameter	Value
User Input	Havayolu Serbestleştirme Kanunu, havacılık endüstrisinde hangi değişiklikleri beraberinde getirdi?
Response	Havayolu Serbestleştirme Kanunu, bilet fiyatları, rotalar ve yeni havayollarının pazara girişi üzerindeki devlet kontrolünü kaldırdı (1).
Retrieved Contexts	Airline Deregulation Act, havacılık endüstrisinde yolcuları piyasa güçlerine maruz bırakmıştır. Airline Deregulation Act, bilet fiyatları, yollar ve yeni havayollarının pazara girişi üzerindeki devlet kontrolünü kaldırmayı amaçlamaktadır. Airline Deregulation Act, 1978'de Amerika Birleşik Devletleri başkanı Jimmy Carter tarafından imzalanmıştır. Enternasyonal turuncu, havacılık endüstrisinde nesneleri çevresinden ayırt etmek için kullanılan bir renktir. Türk Kadastro Kanunu, kadastroyu tanımlar.
Faithfulness	0.0

Contribution of Researchers

Erodoğan BIKMAZ: Implementation, software, evaluation, and writing.

Mohammed BRIMAN Implementation, software, evaluation, and writing.

Serdar ARSLAN: Review and editing.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] F. J. Cazorla *et al.*, "PROXIMA: Improving Measurement-Based Timing Analysis through Randomisation and Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L ., Polosukhin, I.: Attention is All you Need. In: Conference on Neural Information Processing Systems (NeurIPS) (2017). https://papers.nips.cc/paper*files/paper/*2017*/hash/*3*f*5*ee*243547*dee*91*fbd*053*c*1*c*4*a*845*aa*− *Abstract.html*

[2] OpenAI: GPT-4 Technical Report. arXiv preprint (2023). https://arxiv.org/abs/ 2303.08774

[3] Unknown: Introducing Claude 3.5 Sonnet. https://www.anthropic.com/news/ claude-3-5-sonnet. No date provided in citation (n.d.)

[4] Grattafiori: The Llama 3 herd of models. arXiv preprint (2024). https://arxiv. org/abs/2407.21783

[5] Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., De Las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M.A., Stock, P., Scao, T.L., Lavril, T., Wang, T., Lacroix, T., Sayed, W.E.: Mistral 7B. arXiv preprint (2023). https://arxiv.org/abs/2310. 06825

[6] Abdin, M., Aneja, J., Behl, H., Bubeck, S., Eldan, R., Gunasekar, S., Harrison,

M., Hewett, R.J., Javaheripi, M., Kauffmann, P., Lee, J.R., Lee, Y.T., Li, Y., Liu, W., Mendes, C.C.T., Nguyen, A., Price, E., Gustavo, D.R., Saarikivi, O., Salim, A., Shah, S., Wang, X., Ward, R., Wu, Y., Yu, D., Zhang, C., Zhang, Y.: PHI-4 Technical Report. arXiv preprint (2024). https://arxiv.org/abs/2412.08905

[7] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., Zhou, D.: Chain-of-Thought prompting elicits reasoning in large language models. arXiv preprint (2022). https://arxiv.org/abs/2201.11903

[8] Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T.L., Cao, Y., Narasimhan, K.: Tree of Thoughts: Deliberate Problem Solving with Large Language Models. arXiv preprint (2023). https://arxiv.org/abs/2305.10601

[9] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., Liu, T.: A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. ACM Transactions on Office Information Systems (2024) arXiv:2311.05232. Preprint available at https://arxiv.org/abs/2311.05232

[10] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Ku¨ttler, H., Lewis, M., Yih, W.T., Rockt¨aschel, T., Riedel, S., Kiela, D.: Retrieval-Augmented Generation for Knowledge-Intensive NLP tasks. arXiv preprint (2020). https:// arxiv.org/abs/2005.11401

[11] Reimers, N., Gurevych, I.: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv preprint (2019). https://arxiv.org/abs/1908.10084

[12] Jegou, H., Douze, M., Johnson, J.: Faiss: A library for efficient similarity search. Facebook Engineering blog. Published March 29, 2017, but cited as 2018 (2018). https://engineering.fb.com/2017/03/29/data-infrastructure/ faiss-a-library-for-efficient-similarity-search/

[13] Unknown: What is Semantic Search? https://cohere.com/llmu/ what-is-semantic-search. No date provided in citation (n.d.)

[14] Ahmad, S.R.: Enhancing multilingual information retrieval in mixed human resources environments: a RAG model implementation for multicultural enterprise. arXiv preprint (2024). https://arxiv.org/abs/2401.01511

[15] Ustu¨n, A., Aryabumi, V., Yong, Z.X., Ko, W.Y., D'souza, D., Onilude, G., Bhan-¨ dari, N., Singh, S., Ooi, H.L., Kayid, A., Vargus, F., Blunsom, P., Longpre, S., Muennighoff, N., Fadaee, M., Kreutzer, J., Hooker, S.: Aya model: An Instruction finetuned Open-Access Multilingual language model. arXiv preprint (2024). https://arxiv.org/abs/2402.07827

[16] Yang, A., Yang, B., Hui, B., Zheng, B., Yu, B., Zhou, C., Li, C., Li, C., Liu, D., Huang, F., al.: QWEN2 Technical Report. arXiv preprint (2024). https://arxiv.
org/abs/2407.10671

[17] Chirkova, N., Rau, D., D´ejean, H., Formal, T., Clinchant, S., Nikoulina, V.: Retrieval-augmented generation in multilingual settings. arXiv preprint (2024). https://arxiv.org/abs/2407.01463

[18] Chen, W.H.C.S.Y.W.M.K.Z.X.Z.H..Y.D. T.: Dense X retrieval: What retrieval granularity should we use? arXiv preprint (2023). https://arxiv.org/abs/2312. 06648

[19] Gumma, V., Raghunath, A., Jain, M., Sitaram, S.: HEALTH-PARIKSHA: Assessing RAG models for health chatbots in Real-World multilingual settings. arXiv preprint (2024). https://arxiv.org/abs/2410.13671

[20] Xu, P., Wu, H., Wang, J., Lin, R., Tan, L.: Traditional Chinese Medicine Case Analysis System for High-Level Semantic Abstraction: Optimized with Prompt and RAG. arXiv preprint (2024). https://arxiv.org/abs/2411.15491

[21] Zhang, T., Patil, S.G., Jain, N., Shen, S., Zaharia, M., Stoica, I., Gonzalez, J.E.: RAFT: Adapting Language Model to Domain Specific RAG. arXiv preprint (2024). https://arxiv.org/abs/2403.10131

[22] Zhang, T., Patil, S.G., Jain, N., Shen, S., Zaharia, M., Stoica, I., Gonzalez, J.E.: RAFT: Adapting Language Model to Domain Specific RAG. arXiv preprint. This
uses the same source as the 'raft' entry, labelled 'chunking' in the prompt list. (2024). https://arxiv.org/abs/2403.10131

[23] Chen, X., Cardie, C.: Unsupervised multilingual word embeddings. arXiv preprint (2018). https://arxiv.org/abs/1808.08933

[24] Yang, Y., Cer, D., Ahmad, A., Guo, M., Law, J., Constant, N., Abrego, G.H., Yuan, S., Tar, C., Sung, Y.H., Strope, B., Kurzweil, R.: Multilingual universal sentence encoder for semantic retrieval. arXiv preprint (2019). https://arxiv.org/abs/1907.04307

[25] Muennighoff, N., Tazi, N., Magne, L., Reimers, N.: MTEB: Massive Text Embedding Benchmark. arXiv preprint (2022). https://arxiv.org/abs/2210.07316

[26] Unknown: MTEB: Massive Text Embedding Benchmark. https://huggingface.co/ blog/mteb. No date provided in citation (n.d.)

[27] Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., Zhou, M.: MINILM: Deep
Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. arXiv preprint (2020). https://arxiv.org/abs/2002.10957

[28] Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint (2015). https://arxiv.org/abs/1503.02531

[29] Li, Z., Zhang, X., Zhang, Y., Long, D., Xie, P., Zhang, M.: Towards General Text Embeddings with Multi-stage Contrastive Learning. arXiv preprint (2023). https://arxiv.org/abs/2308.03281

[30] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint (2018). https://arxiv.org/abs/1810.04805

[31] Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., Majumder, R., Wei, F.: Text embeddings by Weakly-Supervised contrastive pre-training. arXiv preprint (2022). https://arxiv.org/abs/2212.03533

[32] Hamed, Z., Mostafa, D., Bruce, C.W., Erik, L.-M., Jaap, K.: From Neural ReRanking to Neural Ranking learning a sparse representation for inverted indexing. ACM Proceedings, 497–506 (2018)

[33] Khattab, O., Zaharia, M.: ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT. arXiv preprint (2020). https://arxiv.
org/abs/2004.12832

[34] De Souza P Moreira, G., Ak, R., Schifferer, B., Xu, M., Osmulski, R., Oldridge, E.: Enhancing QA Text Retrieval with Ranking Models: Benchmarking, fine-tuning and deploying Rerankers for RAG (2024). https://arxiv.org/abs/2409.07691

[35] Kesgin, H.T., Yuce, M.K., Amasyali, M.F.: Developing and evaluating tiny to medium-sized turkish bert models. arXiv preprint arXiv:2307.14134 (2023)

[36] ytu-ce-cosmos/turkish-colbert· Hugging Face. https://huggingface.co/ ytu-ce-cosmos/turkish-colbert

[37] Bajaj, P., Campos, D., Craswell, N., Deng, L., Gao, J., Liu, X., Majumder, R., McNamara, A., Mitra, B., Nguyen, T., Rosenberg, M., Song, X., Stoica, A., Tiwary, S., Wang, T.: MS MARCO: A Human Generated MAchine Reading
COmprehension Dataset (2016). https://arxiv.org/abs/1611.09268

[38] MeTin/WIKIRAG-TR · Datasets at Hugging Face. https://huggingface.co/ datasets/Metin/WikiRAG-TR

[39] Gu¨nther, M., Ong, J., Mohr, I., Abdessalem, A., Abel, T., Akram, M.K., Guzman, S., Mastrapas, G., Sturua, S., Wang, B., Werk, M., Wang, N., Xiao, H.: JINA Embeddings 2: 8192-Token General-Purpose text embeddings for long documents (2023). https://arxiv.org/abs/2310.19923

[40] Robinson, J., Chuang, C.-Y., Sra, S., Jegelka, S.: Contrastive Learning with Hard Negative Samples (2020). https://arxiv.org/abs/2010.04592

[41] Thakur, N., Reimers, N., Ru¨ckl´e, A., Srivastava, A., Gurevych, I.: BEIR: a heterogenous benchmark for zero-shot evaluation of information retrieval models (2021). https://arxiv.org/abs/2104.08663

[42] Es, S., James, J., Espinosa-Anke, L., Schockaert, S.: RAGAS: Automated Evaluation of Retrieval Augmented Generation. arXiv preprint (2023). https://arxiv. org/abs/2309.15217

[43] Lin, C.Y.: ROUGE: a package for automatic evaluation of summaries. In: Proceedings of the Meeting of the Association for Computational Linguistics, Barcelona, Spain, pp. 74–81 (2004)

[44] Zhang, T., Kishore, V., Wu, F., Weinberger, K.Q., Artzi, Y.: BERTScore: Evaluating Text Generation with BERT. arXiv preprint (2019). https://arxiv.org/abs/ 1904.09675

[45] LangChain: Recursively split by character. https://python.langchain.com/v0.1/ docs/modules/data connection/document transformers/recursive text splitter/. No date provided in citation (n.d.)

[46] Gemini Team: Gemini: a family of highly capable multimodal models. arXiv preprint (2023). https://arxiv.org/abs/2312.11805

[47] Team, G. et. al.: Gemma 3 Technical Report (2025). https://arxiv.org/abs/2503.19786

[48] Amati, G.: BM25, pp. 257–260 (2009). https://doi.org/10.1007/978-0-387-39940-9\{ 9 . https://doi.org/10.1007/978-0-387-39940-9₉21

[49] Chroma-Core: GitHub - chroma-core/chroma: the AI-native open-source embedding database. https://github.com/chroma-core/chroma

[50] sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 · Hugging Face. https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2